Fast topology identification from smooth graph signals

Seyed Saman Saboksayr*, Gonzalo Mateos*† and Mujdat Cetin*†
*Dept. of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA
†Goergen Institute for Data Science, University of Rochester, Rochester, NY, USA

***Abstract*—We consider network topology identification under a signal smoothness prior. We address said graph learning problem by developing a fast dual-based proximal gradient (FDPG) algorithm that can handle large-scale graphs efficiently. Preliminary results demonstrate the effectiveness of the proposed method in learning graphs accurately and fast.**

***Index Terms*—Graph signal processing, smooth signals, network topology inference, accelerated gradient methods.**

I. Introduction

In various fields of science and engineering, adopting a network-centric vantage point can be instrumental to extract actionable knowledge from relational datasets. Graph signal processing (GSP) proved to be a suitable tool to this end [1]. However, GSP algorithms necessitate a graph representation of complex structures in data, which may be unavailable and has to be inferred from nodal observations [2], [3], [4], [5], [6], [7].

Consider a network described by a weighted and undirected graph $\mathcal{G}\left(\mathcal{V},\mathcal{E},\mathbf{W}\right)$, where $\mathcal{V}=\{1,\ldots,N\}$ represents the node set of cardinality N, $\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ the set of edges, and $\mathbf{W}\in\mathbb{R}_+^{N\times N}$ is the symmetric adjacency matrix. Next, we instroduce graph signals $\mathbf{x}=[x_1,\ldots,x_N]^\top\in\mathbb{R}^N$ over $\mathcal{G}$, where x_i is the signal value at node $i\in\mathcal{V}$.

Signal smoothness with respect to $\mathcal{G}$. The adjacency matrix $\mathbf{W}$ is the descriptor of the graph structure. Accordingly, the combinatorial graph Laplacian $\mathbf{L}:=\text{diag}\left(\mathbf{d}\right)-\mathbf{W}$, where $\mathbf{d}\in\mathbb{R}^N$ is a vector of nodal degrees, can play a central role in defining a measure of signal variability [8]. The total variation (TV) of the graph signal $\mathbf{x}$ with respect to the Laplacian $\mathbf{L}$ (also known as Dirichlet energy) is defined as the following quadratic form

$$\text{TV}(\mathbf{x}):=\mathbf{x}^\top\mathbf{L}\mathbf{x}=\frac{1}{2}\sum_{i\neq j}W_{ij}\left(x_i-x_j\right)^2. \tag{1}$$

The $\text{TV}(\cdot)$ is a smoothness measure, quantifying how much the graph signal $\mathbf{x}$ changes with respect to $\mathcal{G}$'s topology. Smaller values of $\text{TV}(\cdot)$ are indicative of limited signal variability.

Contributions in context of related prior work. In this paper, we develop an algorithmic framework to identify network topology under smoothness priors. Revisiting the general graph learning framework in [6], we adopt a fast dual proximal gradient (FDPG) method to solve the resulting smoothness-regularized optimization problem. It can be shown that the proposed FDPG method has a convergence guarantee [9]. Recent works on graph learning from observations of smooth signals have developed different approaches to solve the said optimization problem [6], [10], [11], [12]. The primal-dual (PD) techniques have been exploited in [6]. PD methods are known to efficiently handle high-dimensional problems. The convergent proximal-gradient (PG) method is introduced in [11] where is amenable to online scenarios. Moreover, the alternating direction method of multipliers (ADMM) is proposed to solve the graph learning optimization problem [12]. Numerical tests using synthetic data indicate the efficiency and effectiveness of the proposed FDPG algorithm in solving the convex minimization. A longer version of this paper with full algorithmic details and convergence analysis along with publicly-available code can be found in [13].

Work in this paper was supported by the NSF awards CCF-1750428, CCF-1934962 and ECCS-1809356. Author emails: ssaboksa@ur.rochester.edu, gmateosb@ece.rochester.edu, and mujdat.cetin@rochester.edu.

II. Graph learning from smooth signals

Given the data matrix $\mathbf{X}=[\mathbf{x}_1,\ldots,\mathbf{x}_T]\in\mathbb{R}^{N\times T}$, and let $\bar{\mathbf{x}}_i^\top\in\mathbb{R}^{1\times T}$ denote its i-th row collecting those T measurements at vertex i. Following [6] we can establish a link between smoothness and sparsity, namely

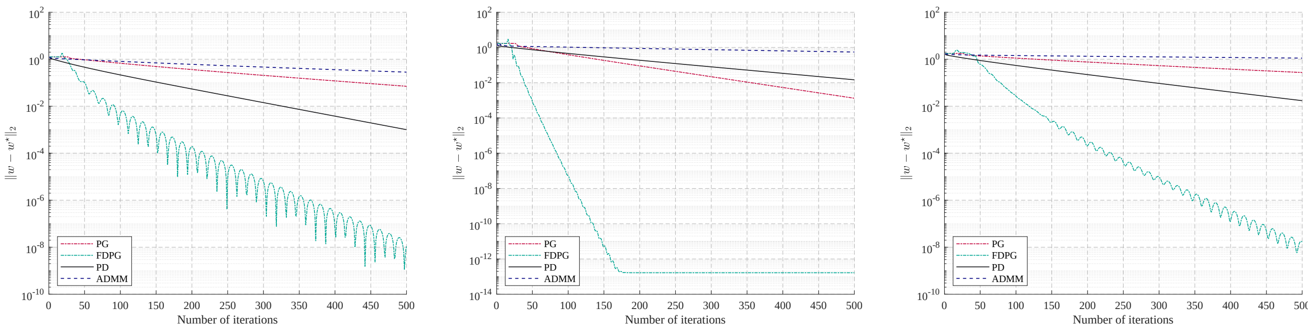
$$\sum_{t=1}^{T}\text{TV}(\mathbf{x}_t)=\text{trace}(\mathbf{X}^\top\mathbf{L}\mathbf{X})=\frac{1}{2}\|\mathbf{W}\circ\mathbf{Z}\|_1, \tag{2}$$

where $\circ$ stands for the Hadamard (element-wise) product and the Euclidean-distance matrix $\mathbf{Z}\in\mathbb{R}_+^{N\times N}$ has entries $Z_{ij}:=\|\bar{\mathbf{x}}_i-\bar{\mathbf{x}}_j\|^2$, $i,j\in\mathcal{V}$. The intuition is that when the given distances in $\mathbf{Z}$ come from a smooth manifold, the corresponding graph has a sparse edge set, with preference given to edges (i,j) associated with smaller distances Z_{ij}.

Leveraging (2) a general graph-learning framework was put forth in [6], which advocates solving the convex smoothness-regularized inverse problem

$$\begin{aligned}\min_{\mathbf{W}}\ &\|\mathbf{W}\circ\mathbf{Z}\|_1-\alpha\mathbf{1}^\top\log(\mathbf{W}\mathbf{1})+\beta\|\mathbf{W}\|_F^2 \\ \text{s. t.}\quad &\text{diag}(\mathbf{W})=\mathbf{0},\ W_{ij}=W_{ji}\geq 0,\ i\neq j.\end{aligned} \tag{3}$$

where $\mathbf{1}$ and $\mathbf{0}$ are vectors of all ones and zeros. Note that $\alpha,\beta>0$ are tuning parameters for controlling the sparsity pattern and scale of the solution [6]. In order to adopt a FDPG method for solving (3), recall first that the adjacency matrix $\mathbf{W}$ is symmetric with diagonal elements equal to zero. Thus, the independent decision variables are effectively the upper-triangular elements $[\mathbf{W}]_{ij}$, $j>i$, which we collect in the vector $\mathbf{w}\in\mathbb{R}_+^{N(N-1)/2}$. Second, it will prove convenient to enforce the non-negativity constraints via a penalty function augmenting the original objective. Just like [6] we



(a) (b) (c)

Fig. 1. Convergence performance on (a) ER graph with 100 nodes, (b) ER graph with 250 nodes, and (c) BA graph with 250 nodes.

add an indicator function $\mathbb{I}\{\mathbf{w} \succeq \mathbf{0}\} = 0$ if $\mathbf{w} \succeq \mathbf{0}$, and $\mathbb{I}\{\mathbf{w} \succeq \mathbf{0}\} = \infty$ otherwise. The superiority performance of (3) has already been shown when compared to other state-of-the-art objective functions [6]. The FDPG method is derived by applying well-known FISTA approach to the dual problem. The FDPG actually does not add any extra computational cost to the problem [9]. The FDPG method can efficiently solve the following minimization problem

$$\min f(\mathbf{x}) + g(\mathbf{S}\mathbf{x}) \tag{4}$$

where $f(\cdot)$ is a strongly convex function with strong convexity parameter σ and $g(\cdot)$ is a convex function [9]. The FDPG method is well-suited for large-scale problems since it enjoys a fast rate of convergence. Interestingly, if we consider the convergence rate of the dual objective function as $\mathcal{O}(1/k^2)$, the primal sequence convergence rate is at $\mathcal{O}(1/k)$ [9].

Given these definitions, we recast the objective in (3) as the function of a vector variable and write the equivalent composite, non-smooth optimization problem

$$\min_{\mathbf{w}} \overbrace{\mathbb{I}\{\mathbf{w} \succeq \mathbf{0}\} + 2\mathbf{w}^\top \mathbf{z} + \beta\|\mathbf{w}\|^2}^{f(\mathbf{w})} \underbrace{-\alpha \mathbf{1}^\top \log(\mathbf{S}\mathbf{w})}_{g(\mathbf{w})}. \tag{5}$$

where $\mathbf{z}$ is a vector containing the upper-triangular entries of $\mathbf{Z}$, and $\mathbf{S} \in \{0,1\}^{N \times N(N-1)/2}$ is such that $\mathbf{d} = \mathbf{W}\mathbf{1} = \mathbf{S}\mathbf{w}$. As a part of FDPG algorithm we have to first compute the following components [9]

$$\underset{\mathbf{x}}{\text{argmax}} \ \langle \mathbf{x}, \mathbf{S}^\top \mathbf{u} \rangle - f(\mathbf{x}) = \max\left(\mathbf{0}, \frac{\mathbf{S}^\top \mathbf{u} - 2\mathbf{z}}{2\beta}\right), \tag{6}$$

$$\mathbf{prox}_{\mu g}(\mathbf{x}) = \frac{\mathbf{x} + \sqrt{\mathbf{x}^2 + 4\alpha\mu}}{2}, \tag{7}$$

where $\max(\cdot)$ in (6) and all operations in (7) are element-wise operations. The resulting iterations based on [9] are tabulated as Algorithm 1. Note that, by choosing a constant step size $\mu = \frac{\|\mathbf{S}\|^2}{\sigma} = \frac{N-1}{\beta}$, the FDPG algorithm is proven to converge; see e.g., [9] and [13] for details.

Algorithm 1: Topology identification via FDPG

Input parameters α, β, μ, initial $\mathbf{u}_1 = \mathbf{y}_0 = \mathbf{0}$, $t_1 = 1$.
for $k = 1, 2, \ldots,$ **do**
  $\mathbf{w}_k = \max\left(\mathbf{0}, \frac{\mathbf{S}^\top \mathbf{u}_k - 2\mathbf{z}}{2\beta}\right)$
  $\mathbf{v}_k = \mathbf{prox}_{\mu g}(\mathbf{S}\mathbf{w}_k - L\mathbf{u}_k)$
  $\mathbf{y}_k = \mathbf{u}_k - \mu^{-1}(\mathbf{S}\mathbf{w}_k - \mathbf{v}_k)$
  $t_{k+1} = 0.5(1 + \sqrt{1 + 4t_k^2})$
  $\mathbf{u}_{k+1} = \mathbf{y}_k + \left(\frac{t_k - 1}{t_{k+1}}\right)(\mathbf{y}_k - \mathbf{y}_{k-1})$
end

III. Preliminary Numerical Results

To assess the performance of the proposed graph learning algorithm, we test it on simulated data. For sake of evaluation, we compare Algorithm 1 to other state-of-the-art methods such as PD [6], PG [11], and ADMM [12]. Throughout, we perform a grid search to determine the best regularization parameters α, β in terms of graph recovery. Also, the ADMM parameters and PD step size are best-tuned for obtaining the best possible convergence rate. We generate three different graphs namely Erdős-Rényi (ER) graphs (edge formation probability $p = 0.2$) with $N = 100$ and $N = 250$ nodes, and Barabási-Albert (BA) graph by adding a new node to the graph each time, connecting to 15 existing nodes in the graph. We simulate 5000 i.i.d. samples that are drawn from a Gaussian distribution $\mathbf{x} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{L}_t^\dagger + \sigma_e^2 \mathbf{I}_N\right)$, where σ_e represents the noise level; see e.g., [7]. As shown in Fig. 1, the proposed method outperforms the other methods in terms of convergence rate.

References

[1] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, 2018.
[2] G. Mateos, S. Segarra, A. G. Marques, and A. Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 16–43, 2019.
[3] X. Dong, D. Thanou, M. Rabbat, and P. Frossard, "Learning graphs from data: A signal representation perspective," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 44–63, 2019.
[4] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 3, pp. 467–483, 2017.

[5] R. Shafipour, S. Segarra, A. G. Marques, and G. Mateos, "Identifying the topology of undirected networks from diffused non-stationary graph signals," *IEEE Open J. Signal Process.*, vol. 2, pp. 171–189, 2021.
[6] V. Kalofolias, "How to learn a graph from smooth signals," in *Artif. Intel. and Stat. (AISTATS)*, 2016, pp. 920–929.
[7] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning Laplacian matrix in smooth graph signal representations," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6160–6173, 2016.
[8] D. Zhou and B. Schölkopf, "A regularization framework for learning from graph data," in *Int. Conf. Mach. Learning (ICML)*, 2004.
[9] A. Beck and M. Teboulle, "A fast dual proximal gradient algorithm for convex minimization and applications," *Operations Research Letters*, vol. 42, no. 1, pp. 1–6, 2014.
[10] S. S. Saboksayr, G. Mateos, and M. Cetin, "Online discriminative graph learning from multi-class smooth signals," *Signal Processing*, vol. 186, p. 108101, 2021.
[11] ——, "Online graph learning under smoothness priors," in *European Signal Process. Conf. (EUSIPCO)*, Dublin, Ireland, 2021.
[12] X. Wang, C. Yao, H. Lei, and A. M.-C. So, "An efficient alternating direction method for graph learning from smooth signals," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2021, pp. 5380–5384.
[13] S. S. Saboksayr, G. Mateos, and M. Cetin, "Accelerated graph learning from smooth signals," *IEEE Signal Process. Lett.*, vol. 28, pp. 2192–2196, 2021.