INFERRING SPARSE GRAPHS FROM SMOOTH SIGNALS WITH THEORETICAL GUARANTEES

Michael G. Rabbat

Department of Electrical and Computer Engineering
McGill University, Montréal, Québec, Canada

ABSTRACT

We consider the problem of inferring a graph from signals which are assumed to be smooth over the graph, in the setting where the graph is also assumed to be sparse. We focus on the case where measurements are Gaussian vectors and the graph topology is encoded in the inverse of the covariance matrix. In addition, the weights of the inverse covariance are assumed to be such that the model is attractive—all partial correlations are non-negative. Unlike other approaches which seek to minimize the Laplacian quadratic form or involve solving a log-det program, we study a simple estimator based on soft thresholding. The estimator involves computing only a single eigenvalue decomposition, and so it can easily scale to networks with thousands of vertices. We provide theoretical results on the reconstruction error as a function of the number of observations and problem dimensions for the case where the underlying graph is assumed to be sparse.

1. INTRODUCTION

The field of signal processing over graphs has emerged as a vibrant and active area of research that promises to generalize classical signal processing theory and methods to data that is more naturally modelled as being supported on a graph [1, 2]. A graph or network is a combinatorial structure that encodes relationships between pairs of nodes, and a graph signal is a function assigning a value to each node in the graph.

A fundamental task for graph signal processing is that of inferring the graph topology from direct or indirect signals. In some applications (e.g., those involving infrastructure networks) the graph may be evident directly from available side information. Even in such cases, the information available may noisy (e.g., available maps may be outdated). In other applications—especially those where the graph captures conceptual relationships (e.g., correlations) rather than physical ones—the graph may need to be inferred from data. In some cases it may desirable to infer the graph so that other graph signal processing methods can subsequently be applied, leveraging knowledge of the graph topology. In other cases the end-goal itself may be to infer the graph topology.

The role of sparsity in covariance matrix estimation was first explored by Dempster [3], who coined the term "covariance selection" for the problem of inferring the positions of non-zero entries in the precision matrix. The past decade has seen significant advances in terms of algorithms for inferring sparse inverse covariance matrices based on convex optimization [4–10]. In many cases these methods are supported by theoretical performance guarantees even in the regime where the number of signals observed, n, is much less than the number of vertices in the graph, p.

Topology inference has also featured prominently in the graph signal processing literature. One line of work (discussed in further detail in Sec. 3) has explored the possibility of inferring graph structure from signals that are assumed to be *smooth* over the graph [11–13]. Another line of work is based on a model where measurements are related to the graph topology via filtering (*e.g.*, modeling the diffusion of an idea or spreading of a disease) [14–17].

This paper considers the problem of inferring a graph encoded in the precision matrix of a Gaussian Markov random field. We focus on the particular case where the precision matrix is directly related to the Laplacian of the unknown graph. We study a simple, closed-form estimator for the precision matrix, and hence the graph Laplacian. Our work is inspired by the elementary estimators proposed in [18]. In the case when the graph is sparse, we prove that the estimator is consistent and provide error bounds.

2. PROBLEM FORMULATION

Let $G = (V, E)$ be an undirected (possibly weighted) graph with $|V| = p$ vertices, $V = \{v_1, \ldots, v_p\}$, and let $\boldsymbol{W} \in \mathbb{R}_+^{p \times p}$ denote its adjacency matrix, with $W_{i,j} > 0$ if and only if $(v_i, v_j) \in E$. The graph Laplacian is the matrix $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$, where $\boldsymbol{D}$ is a diagonal matrix with $D_{i,i} = \sum_{j=1}^{p} W_{i,j}$. For a graph signal $\boldsymbol{x} \in \mathbb{R}^p$, the quadratic form

$$\boldsymbol{x}^\mathsf{T} \boldsymbol{L} \boldsymbol{x} = \sum_{i,j=1}^{p} W_{i,j}(x_i - x_j)^2$$

measures smoothness with respect to the graph topology [1]. The graph Laplacian $\boldsymbol{L}$ is positive semi-definite [19], so $\boldsymbol{x}^\mathsf{T}\boldsymbol{L}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x}$, and $\boldsymbol{x}^\mathsf{T}\boldsymbol{L}\boldsymbol{x}$ is smaller when nodes at neighboring vertices have similar values. In particular, if $\boldsymbol{x} = \mathbf{1}$ then $\boldsymbol{x}^\mathsf{T}\boldsymbol{L}\boldsymbol{x} = 0$.

Let X denote a p-dimensional Gaussian random vector with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$; *i.e.*, $X \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. This distribution is said to be a *Gaussian-Markov random field* (GMRF) with respect to a graph $G = (V, E)$ if the precision matrix, $\boldsymbol{\Theta} = (\boldsymbol{\Sigma})^{-1}$, satisfies $\boldsymbol{\Theta}^\star_{i,j} \neq 0$ if and only if $(v_i, v_j) \in E$. In general, the non-zero off-diagonal elements of $\boldsymbol{\Theta}$ may be positive or negative. When the off-diagonal elements of $\boldsymbol{\Theta}$ are strictly non-positive, the corresponding GMRF is called *attractive* [9, 20], and the partial correlations, which are given by

$$\mathrm{Corr}(X_i, X_j | X_{\setminus\{i,j\}}) = \frac{-\Theta_{i,j}}{\sqrt{\Theta_{i,i}\Theta_{j,j}}},$$

are strictly positive.

As a special case, suppose that $\boldsymbol{\Theta} = \boldsymbol{L}$. Then X has pdf $p(\boldsymbol{x}) \propto \exp(-\boldsymbol{x}^\mathsf{T}\boldsymbol{L}\boldsymbol{x})$; *i.e.*, a signal $\boldsymbol{x}$ is more likely under this model if it

is smooth with respect to the graph, in the sense of having a small Laplacian quadratic form.

Since $\boldsymbol{L}$ is not full rank, taking $\boldsymbol{\Theta} = \boldsymbol{L}$ yields an *improper* GMRF [21]. However, a related proper GMRF can be obtained in the following manner [12,21,22]. When G is undirected, $\boldsymbol{L}$ is symmetric and hence it has a real eigendecomposition $\boldsymbol{L} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\mathsf{T}$, where $\boldsymbol{U}$ is an orthogonal matrix of eigenvectors and $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues. Suppose that

$$X = \boldsymbol{U}Y + W,$$

where $Y \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Lambda}^+)$, $W \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$, and with $\boldsymbol{\Lambda}^+$ denoting the pseudo-inverse of $\boldsymbol{\Lambda}$. Then $\boldsymbol{U}Y \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{L}^+)$, and $X \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{L}^+ + \sigma^2\boldsymbol{I})$ is a proper GMRF with respect to the graph $G = (V, E)$.

Now let $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)} \in \mathbb{R}^p$ denote n i.i.d. realizations of the model just described, and let us organize these into the data matrix $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ with one row per realization.

Problem 1. *Given $\boldsymbol{X}$, we would like to estimate $\boldsymbol{L}$ and hence the structure of the graph $G = (V, E)$ underlying the attractive GMRF.*

Of particular interest in contemporary applications is the case where p is large (many vertices in the graph) and the number of observations n is less than p.

3. MAXIMUM LIKELIHOOD AND ALTERNATIVES

For a general GMRF, given $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Theta}^{-1})$, the maximum likelihood estimate of $\boldsymbol{\Theta}$ is given by the solution to the problem,

$$\begin{array}{ll}\text{minimize} & -\log\det(\boldsymbol{\Theta}) + \operatorname{tr}(\boldsymbol{\Theta}\boldsymbol{S}) \\ \text{subject to} & \boldsymbol{\Theta} \in \mathcal{S}^p_{++}\end{array}$$

where $\mathcal{S}^p_{++}$ is the set of $p \times p$ symmetric positive definite matrices, and $\boldsymbol{S} = \frac{1}{n}\boldsymbol{X}^\mathsf{T}\boldsymbol{X}$ is the sample covariance matrix. This is a convex (semi-definite) program that is challenging to solve in practice for large problem instances (large p) because of the $\log\det$ term.

Note that $\operatorname{tr}(\boldsymbol{\Theta}\boldsymbol{S}) = \frac{1}{n}\operatorname{tr}(\boldsymbol{X}\boldsymbol{\Theta}\boldsymbol{X}^\mathsf{T}) = \frac{1}{n}\sum_{k=1}^n (\boldsymbol{x}^{(k)})^\mathsf{T}\boldsymbol{\Theta}\boldsymbol{x}^{(k)}$. So the maximum likelihood formulation can indeed be viewed as finding a $\boldsymbol{\Theta}$ such that the signals $\boldsymbol{x}^{(k)}$ are smooth with respect to $\boldsymbol{\Theta}$.

Let $\mathcal{L}^p$ denote the set of valid $p \times p$ graph Laplacian matrices,

$$\mathcal{L}^p = \{\boldsymbol{L} \in \mathbb{R}^{p\times p} : L_{i,j} = L_{j,i} \le 0, i \ne j, \text{ and } L_{i,i} = -\sum_{j\ne i} L_{i,j}\}.$$

Under the attractive GMRF model described in the previous section, the maximum likelihood estimate for $\boldsymbol{L}$ can be obtained by solving the problem,

$$\begin{array}{ll}\text{minimize} & -\log\det(\boldsymbol{\Theta}) + \operatorname{tr}(\boldsymbol{\Theta}\boldsymbol{S}) \\ \text{subject to} & \boldsymbol{\Theta}^{-1} = \boldsymbol{L}^+ + \sigma^2\boldsymbol{I} \\ & \boldsymbol{L} \in \mathcal{L}^p, \sigma^2 > 0\end{array}$$

where the minimization is performed jointly over $\boldsymbol{L}$ and σ^2. Again, this is a challenging problem in practice when p is large because of the $\log\det$ term.

Assuming that σ^2 is known, Dong *et al.* [12] propose an alternative formulation,

$$\begin{array}{ll}\text{minimize} & \|\boldsymbol{X} - \boldsymbol{Y}\|_F^2 + \operatorname{tr}(\boldsymbol{Y}\boldsymbol{L}\boldsymbol{Y}^\mathsf{T}) + \alpha\|\boldsymbol{L}\|_F^2 \\ \text{subject to} & \boldsymbol{L} \in \mathcal{L}^p\end{array} \tag{1}$$

where the minimization is over $\boldsymbol{L}$ and $\boldsymbol{Y} \in \mathbb{R}^{n\times p}$. This problem can be viewed as seeking to find $\boldsymbol{Y}$ and $\boldsymbol{L}$ such that $\boldsymbol{Y}$ is simultaneously close to the observations $\boldsymbol{X}$ and smooth with respect to the graph Laplacian $\boldsymbol{L}$. The third term acts as a regularizer on $\boldsymbol{L}$ to control its sparsity, with α being a regularization parameter.

Kalofolias [13] proposes to optimize directly for the weights $\boldsymbol{W}$ rather than for a Laplacian matrix, leading to the problem

$$\begin{array}{ll}\text{minimize} & \|\boldsymbol{W} \odot \boldsymbol{Z}\|_{1,1} + \alpha\boldsymbol{1}^\mathsf{T}\log(\boldsymbol{W}\boldsymbol{1}) + \beta\|\boldsymbol{W}\|_F^2 \\ \text{subject to} & \boldsymbol{W} = \boldsymbol{W}^\mathsf{T} \ge 0, \operatorname{diag}(\boldsymbol{W}) = \boldsymbol{0}\end{array} \tag{2}$$

where $\boldsymbol{Z}$ is a $p\times p$ matrix of pairwise distances, $Z_{i,j} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2$, where $\boldsymbol{x}_i$ denotes the ith column of the data matrix $\boldsymbol{X}$. Here the first term corresponds to $\operatorname{tr}(\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^\mathsf{T})$ in the Laplacian formulation, favouring graphs with respect to which the observed signals are smooth, the logarithmic term controls the degrees of each vertex to ensure the resulting graph estimate is not empty (no edges), and the final Frobenius-norm term controls the sparsity of the solution. Similar to above, α and β are regularization parameters. Kalofolias [13] also proposes a computationally efficient method for solving both (1) and (2) that exhibits very promising performance in experiments.

To date, although these alternative approaches are computationally efficient and scalable, there are no statistical guarantees in the literature for inferring the precision matrix of an attractive GMRF, in particular in the regime where $n < p$.

4. A CLOSED-FORM ESTIMATOR

Inspired by the work of Yang *et al.* [18], in this section we describe an alternative approach to estimating the precision matrix of an attractive GMRF following the model described in Sec. 2.

Suppose we had a symmetric matrix $\boldsymbol{B}(\boldsymbol{X})$ that was close to $\boldsymbol{L}$. Then we could obtain an estimate $\widehat{\boldsymbol{L}}$ by solving the problem

$$\begin{array}{ll}\text{minimize} & \|\boldsymbol{L}\|_{1,\text{off}} \\ \text{subject to} & \|\boldsymbol{L} - \boldsymbol{B}(\boldsymbol{X})\|_{\infty,\text{off}} \le \gamma \\ & L_{i,j} \le 0, \text{ for all } i \ne j \\ & L_{i,i} = -\sum_{j\ne i}^p L_{i,j}, \text{ for } i = 1, \ldots, p\end{array} \tag{3}$$

where $\|\cdot\|_{1,\text{off}}$ and $\|\cdot\|_{\infty,\text{off}}$ refer to the element-wise ℓ_1 and ℓ_∞ norms of the off-diagonal elements of their arguments. The objective can be seen as promoting sparsity of the off-diagonal entries of $\boldsymbol{L}$. The first constraint requires that each off-diagonal element of $\boldsymbol{L}$ be no more than γ away from its counterpart in $\boldsymbol{B}(\boldsymbol{X})$. The second constraint ensures that the off-diagonal entries are non-positive as required for attractive GMRFs. The final constraint ensures that each diagonal entry of $\boldsymbol{L}$ is equal to the sum of the magnitudes of the off-diagonal entries in the same row, as required to be a valid graph Laplacian.

By observing that the problem (3) decomposes into independent convex subproblems for each off-diagonal element, we find that the solution is given in closed form via soft thresholding of the off-diagonal elements. In particular, let

$$T_\gamma(x) = \operatorname{sign}(x)\max(|x| - \gamma, 0)$$

denote the soft-thresholding function, and let $\widetilde{\boldsymbol{B}} = T_\gamma(\boldsymbol{B}(\boldsymbol{X}))$, with T_γ applied element-wise. Then, for $i \ne j$,

$$\widehat{L}_{i,j} = \begin{cases}\widetilde{B}_{i,j} & \text{if } \widetilde{B}_{i,j} \le 0 \\ 0 & \text{otherwise,}\end{cases} \tag{4}$$

and $\widehat{L}_{i,i} = -\sum_{j \neq i} \widehat{L}_{i,j}$. Also note that the problem (3) is only feasible if γ is sufficiently large to allow for nulling any positive off-diagonal entries of $\boldsymbol{B}(\boldsymbol{X})$; *i.e.*, if $\gamma \geq \max_{i \neq j} [\boldsymbol{B}(\boldsymbol{X})]_{i,j}$.

The discussion above is predicated on first finding a suitable statistic $\boldsymbol{B}(\boldsymbol{X})$ that is close to the precision matrix $\boldsymbol{L}$. A natural choice would be to take $\boldsymbol{B}(\boldsymbol{X}) = (\boldsymbol{S} - \sigma^2 \boldsymbol{I})^+$, substituting the sample covariance matrix for $\boldsymbol{\Sigma}$ and supposing that σ^2 is known; alternatively, σ^2 could be estimated by examining the smallest eigenvalue of $\boldsymbol{S}$. However, when $n < p$ then $\boldsymbol{S}$ is rank deficient. The approach suggested in [8, 18] is to first apply an appropriate thresholding operation to $\boldsymbol{S}$ in order to obtain a matrix that is invertible, which turns out to be possible under certain conditions related to the sparsity of $\boldsymbol{L}$ (see Sec. 5).

Let us summarize the procedure for estimating $\boldsymbol{L}$ given $\boldsymbol{X}$. The procedure involves two thresholds $\nu > 0$ and $\gamma > 0$ whose values are to be determined below (see Sec. 5). Then perform the following three steps:

1. Compute $\boldsymbol{B} = (T_\nu(\boldsymbol{S}) - \sigma^2 \boldsymbol{I})^+$,
2. Set $\widetilde{\boldsymbol{B}} = T_\gamma(\boldsymbol{B})$,
3. Set the entries $L_{i,j}$ according to (4), and finally
4. Set the diagonal entries $L_{i,i} = -\sum_{j \neq i} L_{i,j}$.

The main computational effort is the (single) matrix pseudo-inverse calculation performed in the first step, and hence this approach comfortably scales to graphs with thousands of vertices (see Sec. 6).

5. THEORETICAL GUARANTEES

Next we present theoretical guarantees for the estimator described in the previous section. The proof is omitted due to space constraints, but will be made available in an extended version of this paper [23]. The analysis builds on the results of [18], and we comment on the main differences below.

The theoretical guarantees are based on the following assumptions.

A 1. *The Laplacian $\boldsymbol{L}$ has exactly k non-zero off-diagonal entries.*

A 2. *The Laplacian $\boldsymbol{L}$ has bounded ℓ_∞ norm: there exists a finite positive constant κ such that*

$$\|\boldsymbol{L}\|_\infty = \min_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\|\boldsymbol{L}\boldsymbol{x}\|_\infty}{\|\boldsymbol{x}\|_\infty} \leq \kappa.$$

A 3. *The true covariance matrix $\boldsymbol{\Sigma}$ satisfies the conditions that: (i) there exists a positive constant D such that $\max_i \Sigma_{i,i} \leq D$, and (ii) for some $0 \leq q < 1$ there is a positive constant C_q such that $\max_i \sum_{j=1}^p |\Sigma_{i,j}|^q \leq C_q$.*

Assumption A1 is fairly standard, and many graphs of interest in applications are sparse. Assumption A2 is related to the maximum (weighted) degree in the graph; in particular κ is exactly twice the maximum weighted degree. Assumption A3 may appear to be the most restrictive at first glance, requiring that the covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{L}^+ + \sigma^2 \boldsymbol{I}$ be roughly sparse (if $q = 0$ then the condition implies that each row of $\boldsymbol{\Sigma}$ is C_q-sparse). However, it is typically the case that the covariance matrix will be approximately sparse. For an example, see Fig. 1.

With these assumptions in place we can state the main result.

Proposition 1. *Let Assumptions A1–A3 hold, let $a = 16\sqrt{10}(\max_i \Sigma_{i,i})$ and let $p' = \max(p, n)$. Given i.i.d. observations $\boldsymbol{x}^{(k)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \boldsymbol{L}^+ + \sigma^2 \boldsymbol{I}$, let $\widehat{\boldsymbol{L}}$ be the output of the estimator described in Section 4 with parameters $\nu = a\sqrt{\frac{\log p'}{n}}$ and $\gamma = \frac{4a\kappa}{\sigma^2}\sqrt{\frac{\log p'}{n}}$. If $n > C_1 \log p'$, then the bounds*

$$\left\|\boldsymbol{L} - \widehat{\boldsymbol{L}}\right\|_{\infty,off} \leq \frac{8a\kappa}{\sigma^2}\sqrt{\frac{\log p'}{n}}$$

and

$$\left\|\boldsymbol{L} - \widehat{\boldsymbol{L}}\right\|_{1,off} \leq \frac{32a\kappa}{\sigma^2} k \sqrt{\frac{\log p'}{n}},$$

both hold with probability at least $1 - C_2 \exp(-C_3 \log p')$.

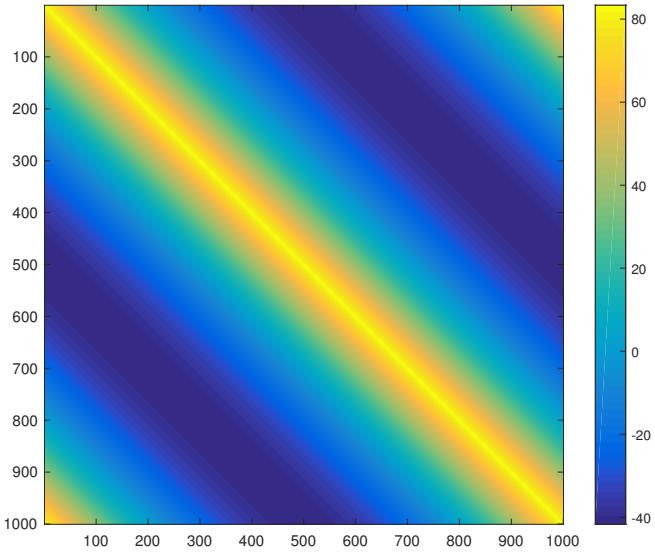

Fig. 1. An example covariance matrix computed as $\boldsymbol{\Sigma} = \boldsymbol{L}^+ + \sigma^2 \boldsymbol{I}$ where the graph used to generate $\boldsymbol{L}$ is a ring on $p = 1000$ vertices and $\sigma^2 = 1e-4$.

The constants C_1, C_2, and C_3 depend on D and C_q from A3. The proof follows similar steps to those of Corollary 1 in [18], with a few key differences. The main difference is that [18] considers the case where the precision matrix $\boldsymbol{\Theta}$ is sparse. However, for the model considered in this paper, this is no longer the case when $\sigma^2 > 0$. This requires adapting the formulation to optimize for $\boldsymbol{L}$ rather than $\boldsymbol{\Theta}$. Hence the algorithm and analysis need to be adapted accordingly. The crux of the analysis involves showing that $\boldsymbol{B}$ computed in the first step is sufficiently close to $\boldsymbol{L}$ for the method to succeed.

6. EXPERIMENTS

We illustrate the method by applying it to synthetic data. We simulate two families of graphs, small-world random graphs following the Watts-Strogatz model [24], and random geometric graphs. For the latter, the vertices are assigned coordinates uniformly and independently in the unit square, and two vertices are connected with an edge if their Euclidean distance is at most $\sqrt{\log(p)/p}$. In both cases we use $\sigma^2 = 0.0001$ and we repeat each experiment 50 times. For the small-world graphs we use $p = 1000$ vertices and $n = 500$ observations, and for the random geometric graphs we use $p = 3000$ vertices and $n = 1500$ observations. In the small-world experiment each node initially has 10 neighbors and links are rewired with probability 0.1. For the parameters used in the random geometric graph experiment, the network has roughly 36300 edges, on average.

The parameters and results are summarized in Table 1. We report the edge accuracy in terms of the average *true positive rate* (TPR) and *false positive rate* (FPR), as well as the average runtime on a contemporary laptop with 16GB of RAM and an 2.9GHz Intel Core i5 processor. Although the performance is modest, it is notable that the approach indeed offers interesting results, especially given the simplicity of the estimator. The runtime is dominated by the pseudo-inverse calculation, which we implement via an eigenvalue decomposition. Even for a graph with 3000 vertices the procedure terminates in well under two seconds. Additional improvement in performance may be possible with additional fine-tuning of the threshold parameters.

	n	p	TPR	FPR	time (s)
Small world	500	1000	0.9256	0.0636	0.0756
Rand. geo.	1500	3000	0.9557	0.2142	1.1873

Table 1. Experiment parameters and results.

7. DISCUSSION

This paper studied a simple threshold-based estimator for inferring the topology of a sparse graph from signals that are assumed to be smooth over the graph. The inference procedure is computationally simple, involving three steps. The computational complexity is dominated by one matrix inverse. The other operations amount to soft thresholding the sample covariance matrix and an approximate Laplacian. Theoretical guarantees are provided, and experiments illustrate the performance on networks with up to thousands of vertices.

The approach considered in this paper is fundamentally based on the assumption that the network is sparse, rather than that the signals are smooth with respect to the graph. Indeed, the majority of the theoretical guarantees available in the literature for covariance selection of large graphical models (including general Gaussian graphical models) are based on sparsity assumptions. Continuing to expand this literature on topology inference methods which are scalable and which also have performance guarantees is of significant interest.

More recently in the graph signal processing literature there has been equal or more emphasis on finding graphs for which the observed signals are *smooth*, with sparsity of the topology being secondary. One may ask if smoothness (*e.g.*, as measured by the Laplacian quadratic form) is sufficient to provide topology recovery guarantees, without any requirement for sparsity. While it is interesting to pursue this question further, we conjecture that the answer is negative. The Laplacian quadratic form $\boldsymbol{x}^\mathsf{T}\boldsymbol{L}\boldsymbol{x} = \sum_{j=1}^{p} \lambda_j(\boldsymbol{L}) \left\|\boldsymbol{u}_j^\mathsf{T}\boldsymbol{x}\right\|_2^2$ can be interpreted as a weighted sum of the energy in each subspace spanned by an eigenvector, scaled by the corresponding eigenvalue. The existence of spectral sparsifiers [25]—a pair of graphs, G and $\widehat{G}$, one of which is dense and the other sparse, both of which have approximately the same eigenvalues—suggests that methods seeking to find graphs only based on smoothness will, in general, have multiple solutions. Regularizing for sparsity is an obvious way to arrive at a unique solution.

8. REFERENCES

[1] D. 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*, vol. 30, no. 3, pp. 83–98, May 2013.

[2] A. Sandryhaila and J.M.F. Moura, "Big data analysis with signal processing on graphs," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 80–90, Sep. 2014.

[3] A.P. Dempster, "Covariance selection," *Biometrics*, vol. 28, no. 1, pp. 157–175, Mar. 1972.

[4] N. Meinshausen and P. Bühlmann, "High-dimensional graphs and variable selection with the Lasso," *Annals of Statistics*, vol. 34, no. 3, pp. 1436–1462, 2006.

[5] P. Ravikumar, M.J. Wainwright, and J.D. Lafferty, "High-dimensional Ising model selectin using ℓ_1-regularized logistic regression," *Annals of Statistics*, vol. 38, no. 3, pp. 1287–1319, 2010.

[6] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical Lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, Jul. 2008.

[7] O. Banerjee, L. El Ghaoui, and A. d'Aspremont, "Model selection through sparse maximum likleihood estimation for multivariate Gaussian or binary data," *Journal of Machine Learning Research*, vol. 9, pp. 485–516, Mar. 2008.

[8] A.J. Rothman, E. Levina, and J. Zhu, "Generalized thresholding of large covariance matrices," *J. American Statistical Association*, vol. 104, no. 485, pp. 177–186, Mar. 2009.

[9] A. Anandkumar, V.Y.F. Tan, F. Huang, and A.S. WIllsky, "High-dimensional Gaussian graphical model selection: Walk summability and local separation criterion," *Journal of Machine Learning Research*, vol. 13, pp. 2293–2337, Aug. 2012.

[10] G. Marjanovic and A.O. Hero, "ℓ_0 sparse inverse covariance estimation," *IEEE Trans. Signal Processing*, vol. 63, no. 12, pp. 3218–3231, May 2015.

[11] E. Pavez and A. Ortega, "Generalized Laplacian precision matrix estimation for graph signal processing," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Shanghai, China, Mar. 2016.

[12] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning Laplacian matrix in smooth graph signal representations," preprint, available onine at `http://arxiv.org/abs/1406.7842`, Feb. 2016.

[13] V. Kalofolias, "How to learn a graph from smooth signals," in *Proc. AISTATS*, Cadiz, Spain, May 2016.

[14] J. Mei and J.M.F. Moura, "Signal processing on graphs: Estimating the structure of a graph," in *Proc. ICASSP*, Brisbane, Australia, Apr. 2015.

[15] B. Pasdeloup, M. Rabbat, V. Gripon, D. Pastor, and G. Mercier, "Graph reconstruction from the observation of diffused signals," in *Proc. Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, USA, Sep. 2015.

[16] S. Segarra, A.G. Marques, G. Mateos, and A. Ribeiro, "Network topology identification from spectral templates," in *Proc. IEEE SSP*, Palma de Mallorca, Spain, Jun. 2016.

[17] Y. Shen, B. Baingana, and G.B. Giannakis, "Nonlinear structural equation models for network topology inference," in *Proc. Conf. on Info. Sciences and Systems*, Princeton, NJ, USA, Mar.

[18] E. Yang, A.C. Lozano, and P. Ravikumar, "Elementary estimators for graphical models," in *Proc. NIPS*, Montreal, Canada, Dec. 2014.

[19] C. Godsil and G. Royle, *Algebraic Graph Theory*, Springer, 2001.

[20] M. Slawski and M. Hein, "Estimation of positive definite M-matrices and structure learning for attractive Gaussian Markov random fields," *Linear Algebra and its Applications*, vol. 473, pp. 145–179, 2015.

[21] C. Zhang, D. Florêncio, and P.A. Chou, "Graph signal processing: A probabilistic framework," Technical Report MSR-TR-2015-31, Microsoft Research, Redmond, USA, Apr. 2015.

[22] B. Lake and J. Tennenbaum, "Discovering structure by learning sparse graphs," in *Proc. Annual Conf. of the Cognitive Science Society*, Portland, OR, USA, Aug. 2010.

[23] M. Rabbat, "Inferring sparse graphs from smooth signals with theoretical guarantees (extended version)," Tech. report, McGill University, Sep. 2016.

[24] D.J. Watts and S.H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[25] J. Batson, D. Spielman, N. Srivastava, and S.H. Teng, "Spectral sparsification of graphs: Theory and algorithms," *Communications of the ACM*, vol. 56, no. 8, pp. 87–94, 2013.