2016 IEEE 55th Conference on Decision and Control (CDC)
ARIA Resort & Casino
December 12-14, 2016, Las Vegas, USA

Topology identification of undirected consensus networks via sparse inverse covariance estimation

Sepideh Hassan-Moghaddam, Neil K. Dhingra, and Mihailo R. Jovanović

***Abstract*— We study the problem of identifying sparse interaction topology using sample covariance matrix of the state of the network. Specifically, we assume that the statistics are generated by a stochastically-forced undirected first-order consensus network with unknown topology. We propose a method for identifying the topology using a regularized Gaussian maximum likelihood framework where the ℓ_1 regularizer is introduced as a means for inducing sparse network topology. The proposed algorithm employs a sequential quadratic approximation in which the Newton's direction is obtained using coordinate descent method. We provide several examples to demonstrate good practical performance of the method.**

***Index Terms*— Convex optimization, coordinate descent, ℓ_1 regularization, maximum likelihood, Newton's method, sparse inverse covariance estimation, sequential quadratic approximation, topology identification, undirected consensus networks.**

I. Introduction

Recovering network topology from partially available statistical signatures is a challenging problem. Due to experimental or numerical limitations, it is often the case that only noisy partial network statistics are known. Thus, the objective is to account for all observed correlations that may be available and to develop efficient convex optimization algorithms (for network topology identification) that are well-suited for big data problems.

Over the last decade, a rich body of literature has been devoted to the problems of designing network topology to improve performance [1]–[8] and identifying an unknown network topology from available data [9]–[16]. Moreover, the problem of sparsifying a network with dense interconnection structure in order to optimize a specific performance measure has been studied in [17], [18]. In [19], it was demonstrated that individual starlings within a large flock interact only with a limited number of immediate neighbors. The authors have shown that the networks with six or seven nearest neighbors provide an optimal trade-off between flock cohesion and effort of individual birds. These observations were made by examining how robustness of the group depends on the network structure. The proposed measure of robustness captures the network coherence [20] and it is well-suited for developing tractable optimization framework.

Financial support from the 3M Graduate Fellowship, from the University of Minnesota Doctoral Dissertation Fellowship, the University of Minnesota Informatics Institute Transdisciplinary Faculty Fellowship, and the National Science Foundation under award ECCS-1407958 is gratefully acknowledged.

Sepideh Hassan-Moghaddam, Neil K. Dhingra, and Mihailo R. Jovanović are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455. E-mails: hassa247@umn.edu, dhin0008@umn.edu, mihailo@umn.edu.

In this paper, we develop a convex optimization framework for identifying sparse interaction topology using sample covariance matrix of the state of the network. Our framework utilizes an ℓ_1-regularized Gaussian maximum likelihood estimator. Because of strong theoretical guarantees, this approach has been commonly used for recovering sparse inverse covariance matrices [21]–[25]. We utilize the structure of undirected networks to develop an efficient second-order method based on a sequential quadratic approximation. As in [26], [27], we compute the Newton's direction using coordinate descent method [28]–[30] that employs active set strategy. The main point of departure is the formulation of a convex optimization problem that respects particular structure of undirected consensus networks. We also use a reweighted ℓ_1-heuristics as an effective means for approximating non-convex cardinality function [31], thereby improving performance relative to standard ℓ_1 regularization.

Our presentation is organized as follows. In Section II, we formulate the problem of topology identification using sparse inverse covariance matrix estimation. In Section III, we develop a customized second-order algorithm to solve the ℓ_1-regularized Gaussian maximum likelihood estimation problem. In Section IV, we use computational experiments to illustrate features of the method. In Section V, we conclude with a brief summary.

II. Problem formulation

In this section, we provide background material on stochastically forced undirected first-order consensus networks and formulate the problem of topology identification using a sample covariance matrix. The inverse of a given sample covariance matrix can be estimated using Gaussian maximum likelihood estimator. For undirected consensus networks, we show that the estimated matrix is related to the graph Laplacian of the underlying network.

Our objective is to identify the underlying graph structure of a stochastically forced undirected consensus network with a known number of nodes by sampling its second-order statistics. In what follows, we relate the problem of topology identification for consensus networks to the inverse covariance matrix estimation problem.

A. Undirected consensus networks

We consider an undirected consensus network

$$\dot{\psi} \;=\; -L_x\,\psi \,+\, d, \tag{1}$$

where $\psi \in \mathbb{R}^n$ represents the state of n nodes, d is the disturbance input, and the symmetric $n \times n$ matrix L_x

represents the graph Laplacian. The matrix L_x is restricted to have an eigenvalue at zero corresponding to an eigenvector of all ones, $L_x \mathbb{1} = 0$. This requirement is implicitly satisfied by expressing L_x in terms of the incidence matrix E

$$L_x \;:=\; \sum_{l=1}^{m} x_l\, \xi_l\, \xi_l^T \;=\; E\, \mathrm{diag}\,(x)\, E^T,$$

where $\mathrm{diag}\,(x)$ is a diagonal matrix containing the vector of edge weights $x \in \mathbb{R}^m$. Each column $\xi_l = \mathrm{e}_i - \mathrm{e}_j$, where $\mathrm{e}_i \in \mathbb{R}^n$ is the ith basis vector, represents an edge connecting nodes i and j. The m columns of E specify the edges that may be used to construct the consensus network. For a complete graph, there are $m = n(n-1)/2$ potential edges.

In order to achieve consensus in the absence of disturbances, it is required that the closed-loop graph Laplacian, L_x, be positive definite on the orthogonal complement of the vector of all ones, $\mathbb{1}^\perp$ [32]. This amounts to positive definiteness of the "strengthened" graph Laplacian of the closed-loop network

$$X \;:=\; L_x \;+\; (1/n)\, \mathbb{1}\mathbb{1}^T \;\succ\; 0. \tag{2}$$

Clearly, since $L_x \mathbb{1} = 0$, $X\mathbb{1} = \mathbb{1}$.

Consensus networks attempt to compute the network average; thus, it is of interest to study the deviations from average

$$\tilde{\psi}(t) \;:=\; \psi(t) \;-\; \mathbb{1}\, \bar{\psi}(t) \;=\; \left(I \;-\; (1/n)\, \mathbb{1}\mathbb{1}^T\right) \psi(t),$$

where $\bar{\psi}(t) := (1/n)\mathbb{1}^T\psi(t)$ is the network average and corresponds to the zero eigenvalue of L_x. From (1), it follows that the dynamics of the deviation from average are,

$$\dot{\tilde{\psi}} \;=\; -\, L_x \tilde{\psi} \;+\; \left(I \;-\; (1/n)\, \mathbb{1}\mathbb{1}^T\right) d.$$

The steady-state covariance of $\tilde{\psi}$,

$$P \;:=\; \lim_{t\,\to\,\infty} \mathbf{E}\left(\tilde{\psi}(t)\, \tilde{\psi}^T(t)\right),$$

is given by the solution to the algebraic Lyapunov equation

$$L_x\, P \;+\; P\, L_x \;=\; I \;-\; (1/n)\, \mathbb{1}\mathbb{1}^T.$$

For connected networks, the unique solution is given by

$$\begin{aligned} P \;=\; \frac{1}{2} L_x^\dagger \;&=\; \frac{1}{2}\left(\left(L_x \;+\; (1/n)\, \mathbb{1}\mathbb{1}^T\right)^{-1} \;-\; (1/n)\, \mathbb{1}\mathbb{1}^T\right) \\ &=\; \frac{1}{2}\left(X^{-1} \;-\; (1/n)\, \mathbb{1}\mathbb{1}^T\right), \end{aligned} \tag{3}$$

where $(\cdot)^\dagger$ is the pseudo-inverse of a matrix. Thus, the inverse of the steady-state covariance matrix of the deviation from network average is determined by the strengthened graph Laplacian of the consensus network X.

B. Topology identification

A sparse precision matrix can be obtained as the solution to the regularized maximum log-likelihood problem [23],

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & -\log\det\,(X) \;+\; \mathrm{trace}\,(S\, X) \;+\; \gamma\, \|X\|_1 \\ \text{subject to} \quad & X \;\succ\; 0, \end{aligned} \tag{4}$$

where S is the sample covariance matrix, γ is a positive regularization parameter, and $\|X\|_1 := \sum |X_{ij}|$ is the ℓ_1 norm of the matrix X. The ℓ_1 norm is introduced as a means for inducing sparsity in the inverse covariance matrix where a zero element implies conditional independence. This problem has received significant attention in recent years [21], [23], [26], [27], [33]–[36].

In this work, we establish a relation between inverse covariance matrix estimation and the problem of topology identification of an undirected consensus network. We are interested in identifying a sparse topology that yields close approximation of a given sample covariance matrix. This is achieved by solving the following problem,

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & J(x) \;+\; \gamma \sum_{l=1}^{m} |x_l| \\ \text{subject to} \quad & E\, \mathrm{diag}\,(x)\, E^T \;+\; (1/n)\, \mathbb{1}\mathbb{1}^T \;\succ\; 0, \end{aligned} \tag{NI}$$

where

$$\begin{aligned} J(x) \;=\; & -\log\det\left(E\, \mathrm{diag}\,(x)\, E^T \;+\; (1/n)\, \mathbb{1}\mathbb{1}^T\right) \\ & +\; \mathrm{trace}\,(S\, E\, \mathrm{diag}\,(x)\, E^T). \end{aligned}$$

Relative to [21], [26], [27], [33], [34], our optimization problem has additional structure induced by the dynamics of undirected consensus networks.

The network identification problem (NI) is a convex but non-smooth problem where the optimization variable is the vector of the edge weights $x \in \mathbb{R}^m$ and the problem data is the sample covariance matrix S and the incidence matrix. The incidence matrix is selected to contain all possible edges. The ℓ_1 norm of x is a convex relaxation of the cardinality function and it is introduced to promote sparsity. The positive parameter γ specifies the emphasis on sparsity versus matching the sample covariance matrix S. For $\gamma = 0$, the solution to (NI) is typically given by a vector x with all non-zero elements. The positive definite constraint comes from (2) and guarantees a connected closed-loop network and thus asymptotic consensus in the absence of disturbances.

Remark 1: We also use this framework to address the network sparsification problem where it is of interest to find a sparse network that generates close approximation of the covariance matrix of a given dense network. We choose E to be equal to the incidence matrix of the primary network.

III. Customized algorithm based on sequential quadratic approximation

We next exploit the structure of the optimization problem (NI) to develop an efficient customized algorithm. Our algorithm is based on sequential quadratic approximation of the smooth part J of the objective function in (NI). This method benefits from exploiting second-order information about J and from computing the Newton direction using cyclic coordinate descent [28]–[30] over the set of active variables. We find a step-size that ensures the descent direction via backtracking line search. Furthermore, by restricting our computations to active search directions, computational cost is significantly reduced. A similar approach has been

recently utilized in a number of applications, including sparse inverse covariance estimation in graphical models [26], [27], [37]. In this work, we have additional structural constraints and use reweighted heuristics in order to achieve sparsity. We use an alternative proxy for promoting sparsity which is given by the weighted ℓ_1 norm [31]. In particular, we solve problem (NI) for different values of γ using a path-following iterative reweighted algorithm; see Section II (A) in [38]. The topology identification then is followed by a polishing step to debias the identified edge weights.

A. Structured problem: debiasing step

In addition to promoting sparsity of the identified edge weights, the ℓ_1 norm penalizes the magnitude of the nonzero edge weights. In order to gauge the performance of the estimated network topology, once we identify a set of sparse topology via (NI), we solve the structured "polishing" or "debiasing" problem to optimize J over the set of identified edges. To do this, we form a new incidence matrix $\hat{E}$ which contains only those edges identified as nonzero in the solution to (NI) and form the problem,

$$\begin{array}{ll} \underset{x}{\text{minimize}} & -\log\det\left(\hat{E}\,\mathrm{diag}\,(x)\,\hat{E}^T \,+\, (1/n)\,\mathbb{1}\mathbb{1}^T\right) + \\ & \mathrm{trace}\left(S\,\hat{E}\,\mathrm{diag}\,(x)\,\hat{E}^T\right) \\ \text{subject to} & \hat{E}\,\mathrm{diag}\,(x)\,\hat{E}^T \,+\, (1/n)\,\mathbb{1}\mathbb{1}^T \,\succ\, 0. \end{array}$$

whose solution provides the optimal estimated graph Laplacian with the desired structure.

B. Gradient and Hessian of $J(x)$

We next derive the gradient and Hessian of J which can be used to form a second-order Taylor series approximation of $J(x)$ around x^k,

$$J(x^k+\tilde{x}) \;\approx\; J(x^k) \,+\, \nabla J(x^k)^T\tilde{x} \,+\, \frac{1}{2}\,\tilde{x}^T\,\nabla^2 J(x^k)\,\tilde{x}. \tag{5}$$

Proposition 1: The gradient and the Hessian of J at x^k are

$$\begin{array}{rcl} \nabla J(x^k) &=& \mathrm{diag}\left(E^T(S \,-\, X^{-1}(x^k))\,E\right) \\ \nabla^2 J(x^k) &=& M(x^k)\circ M(x^k) \end{array} \tag{6}$$

where $\circ$ denotes the elementwise (Hadamard) product and

$$\begin{array}{rcl} X^{-1}(x^k) &:=& \left(E\,D_{x^k}\,E^T \,+\, (1/n)\,\mathbb{1}\mathbb{1}^T\right)^{-1}, \\ M(x^k) &:=& E^T X^{-1}(x^k)\,E. \end{array}$$

Proof: Utilizing the second order expansion of the log-determinant function we have

$$J(x^k \,+\, \tilde{x}) \,-\, J(x^k) \,\approx\, \mathrm{trace}\left(E^T(S \,-\, X^{-1}(x^k))\,E\,D_{\tilde{x}}\right)$$
$$+\,\frac{1}{2}\,\mathrm{trace}\left(D_{\tilde{x}}\,E^T X^{-1}(x^k)\,E\,D_{\tilde{x}}\,E^T X^{-1}(x^k)\,E\right).$$

The expressions in (6) can be established using a sequence of straightforward algebraic manipulations in conjunction with the use of the commutativity invariance of the trace function and the following properties for a matrix T, a vector α, and a diagonal matrix D_α,

$$\begin{array}{rcl} \mathrm{trace}\,(T\,D_\alpha) &=& \alpha^T\,\mathrm{diag}\,(T) \\ \mathrm{trace}\,(D_\alpha\,T\,D_\alpha\,T) &=& \alpha^T\,(T\circ T)\,\alpha. \end{array}$$

■

C. Algorithm

Our algorithm is based on building the second-order Taylor series expansion of the smooth part of the objective function J in (NI) around the current iterate x^k. Approximation J in (NI) with (5),

$$\begin{array}{ll} \underset{\tilde{x}}{\text{minimize}} & \nabla J(x^k)^T\,\tilde{x} \,+\, \frac{1}{2}\,\tilde{x}^T\,\nabla^2 J(x^k)\,\tilde{x} \,+\, \gamma\,\|x^k+\tilde{x}\|_1 \\ \text{subject to} & E\,\mathrm{diag}\left(x^k \,+\, \tilde{x}\right)E^T \,+\, (1/n)\,\mathbb{1}\mathbb{1}^T \,\succ\, 0. \end{array} \tag{7}$$

We use the coordinate descent algorithm to determine the Newton direction. Let $\tilde{x}$ denote the current iterate approximating the Newton direction. By perturbing $\tilde{x}$ in the direction of the ith standard basis vector $\mathrm{e}_i \in \mathbb{R}^m$, $\tilde{x} \,+\, \mu_i\,\mathrm{e}_i$, the objective function in (7) becomes

$$\begin{array}{rl} \nabla J(x^k)^T(\tilde{x}+\mu_i\,\mathrm{e}_i) \,+\, \frac{1}{2}\,(\tilde{x}+\mu_i\,\mathrm{e}_i)^T\,\nabla^2 J(x^k)\,(\tilde{x}+\mu_i\,\mathrm{e}_i) & \\ +\,\gamma\,|x_i^k \,+\, \tilde{x}_i \,+\, \mu_i|. & \end{array}$$

Elimination of constant terms allows us to express (7) as

$$\underset{\mu_i}{\text{minimize}} \quad \frac{1}{2}\,a_i\,\mu_i^2 \,+\, b_i\,\mu_i \,+\, \gamma\,|c_i \,+\, \mu_i| \tag{8}$$

where (a_i, b_i, c_i, x_i^k, $\tilde{x}_i$) are the problem data,

$$\begin{array}{rcl} a_i &:=& \mathrm{e}_i^T\,\nabla^2 J(x^k)\,\mathrm{e}_i \\ b_i &:=& \left(\nabla^2 J(x^k)\,\mathrm{e}_i\right)^T\tilde{x} \,+\, \mathrm{e}_i^T\,\nabla J(x^k) \\ c_i &:=& x_i^k \,+\, \tilde{x}_i. \end{array}$$

The explicit solution to (8) is given by

$$\mu_i \;=\; -\,c_i \,+\, \mathcal{S}_{\gamma/a_i}(c_i \,-\, b_i/a_i)\,,$$

where $\mathcal{S}_\kappa(y) \;=\; \mathrm{sign}\,(y)\,\max\,(|y| \,-\, \kappa, 0)$, is the soft-thresholding function.

After the Newton direction $\tilde{x}$ has been computed, we determine the step-size α via backtracking. This guarantees positive definiteness of the strengthened graph Laplacian and sufficient decrease of the objective function. We use Armijo rule to find an appropriate step-size such that $E\,\mathrm{diag}(x^k + \alpha\tilde{x})E^T + (1/n)\mathbb{1}\mathbb{1}^T$ is positive definite matrix and

$$J(x^k+\alpha\tilde{x}) \,+\, \gamma\,\|x^k+\alpha\tilde{x}\|_1 \,\leq\, J(x^k) \,+\, \gamma\,\|x^k\|_1 \,+$$
$$\alpha\,\sigma\left(\nabla J(x^k)^T\tilde{x} \,+\, \gamma\,\|x^k+\alpha\tilde{x}\|_1 \,-\, \gamma\,\|x^k\|_1\right).$$

There are two computational aspects in our work which lead to suitability of this algorithm for large-scale networks.

- **Active set strategy**
  We propose an active set prediction strategy as an efficient method to solve the problem (NI) for large values of γ. It is an effective means for determining which directions need to be updated in the coordinate descent algorithm. The classification of a variable as

either active or inactive is based on the values of x_i^k and the ith component of the gradient vector $\nabla J(x^k)$. Specifically, the ith search direction is only inactive if

$$x_i^k \;=\; 0 \;\;\text{and}\;\; |\,\mathrm{e}_i^T\,\nabla J(x^k)\,| \;<\; \gamma \;-\; \epsilon$$

where $\epsilon > 0$ is a small number (e.g., $\epsilon = 0.0001\gamma$). At each outer iteration, the Newton search direction is obtained by solving the optimization problem over the set of active variables. The size of active sets is small for large values of the regularization parameter γ.

- **Memory saving**
  Computation of b_i in (8) requires a single vector inner product between the ith column of the Hessian and $\tilde{x}$, which typically takes $O(m)$ operations. To avoid direct multiplication, in each iteration after finding μ_i, we update the vector $\nabla^2 J(x^k)^T\tilde{x}$ using the correction term $\mu_i(E^T X^{-1}\xi_i) \circ ((X^{-1}\xi_i)^T E)^T$ and take its ith element to form b_i. Here, ξ_i is the ith column of the incidence matrix of the controller graph. This also avoids the need to store the Hessian of J, which is an $m \times m$ matrix, thereby leading to a significant memory saving. Moreover, the ith column of $\nabla^2 J(x^k)$ and the ith element of the gradient vector $\nabla J(x^k)$ enter into the expression for b_i. On the other hand, a_i is determined by the ith diagonal element of the Hessian matrix $\nabla^2 J(x^k)$. All of these can be obtained directly from $\nabla^2 J(x^k)$ and $\nabla J(x^k)$ which are formed in each outer iteration without any multiplication.

Our problem is closely related to the problem in [27]. The objective function there has the form $f(x) = J(x) + g(x)$, where $J(x)$ is smooth over the positive definite cone, and $g(x)$ is a separable non-differentiable function. In our problem formulation, $J(x)$ is smooth for $E\,\mathrm{diag}(x)\,E^T + (1/n)\mathbb{1}\mathbb{1}^T \succ 0$ while the non-smooth part is the ℓ_1 norm which is separable. Thus, convergence can be established using similar arguments. According to [27, Theorems 1,2], the quadratic approximation method converges to the unique global optimum of (NI) and at super-linear rate.

The optimality condition for any x^* that satisfies $E\,\mathrm{diag}(x^*)\,E^T + (1/n)\mathbb{1}\mathbb{1}^T \succ 0$ is given by

$$\nabla J(x^*) \;+\; \gamma\,\partial\,\|x^*\|_1 \;\in\; 0,$$

where $\partial\|x^*\|_1$ is the subgradient of the ℓ_1 norm. This means that for any i

$$\nabla_i J(x^*) \;\in \begin{cases} -\,\gamma, & x_i > 0; \\ \gamma, & x_i < 0; \\ [-\,\gamma, \gamma], & x_i = 0. \end{cases}$$

The stopping criterion is to check the norm of $\nabla J(x)$ and the sign of x to make sure that x is the optimal solution.

IV. Computational experiments

We next illustrate the performance of our customized algorithm. We have implemented our algorithm in Matlab, and all tests were done on 3.4 GHz Core(TM) i7-3770 Intel(R) machine with 16GB RAM.

The problem (NI) is solved for different values of γ using the path-following iterative reweighted algorithm [31] with $\varepsilon = 10^{-5}$. The initial weights are computed using the solution to (NI) with $\gamma = 0$ (i.e., the optimal centralized vector of the edge weights). We then adjust $\varepsilon = 0.001\|x\|_2$ at each iteration. Topology identification is followed by the polishing step described in Section III-A. In the figures, we use black dots to represent nodes, blue lines to identify the original graph, and red lines to denote the edges in the estimated sparse network. In all examples, we set the tolerance for the stopping criterion to 10^{-4}.

A. Network identification

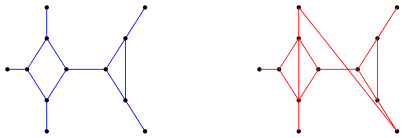
We solve the problem of identification of a sparse network using sample covariance matrix for 500 logarithmically-spaced values of $\gamma \in [0.1, 1000]$. The sample covariance matrix S is obtained by sampling the nodes of the stochastically-forced undirected unweighted network whose topology is shown in Fig. 1a. To generate samples, we conducted 20 simulations of system (1) forced with zero-mean unit variance band-limited white noise d. The sample covariance matrix is averaged over all simulations and asymptotically converges to the steady-state covariance matrix. The incidence matrix E in (NI) contains all possible edges.

Empirically, we observe that after about 5 seconds the sample covariance matrix converges to the steady-state covariance. First, we sample the states after 3 seconds, so the sample covariance matrix we compute is different than the true steady-state covariance matrix. For (NI) solved with this problem data, Figures 3 and 1c illustrate the topology of the identified networks for the minimum and maximum values of γ. In Fig. 1d, the blue line shows an edge in the original network that has not been recovered by the algorithm for the largest value of γ. The red lines in this figure show two extra edges in the estimated network for the smallest value of γ which were not present in the original graph. Next, we solve the problem (NI) using a sample covariance matrix generated by sampling the node states after 15 seconds. In this case, the sample covariance matrix is closer to the steady-state covariance matrix than the previous experiment. As shown in Fig. 2a, the identified network is exactly the same as the original network for $\gamma = 0.1$. If γ is further increased, a network sparser than the original is identified; see Fig. 2b.

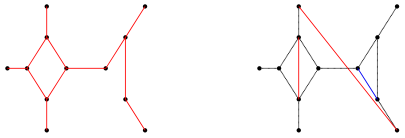
For $\gamma = 0$, the relative error between the covariance matrix of the estimated network and the sample covariance matrix S is given by

$$\frac{\|\left(E\,\mathrm{diag}\,(x_c)\,E^T \;+\; (1/n)\,\mathbb{1}\,\mathbb{1}^T\right)^{-1} \;-\; S\|_F}{\|S\|_F} \;=\; 0.004\%,$$

where $\|\cdot\|_F$ is the Frobenius norm and x_c is the solution to (NI) with $\gamma = 0$. As γ increases, the number of nonzero elements in the vector of the edge weights x decreases and the state covariance matrix gets farther away from the given sample covariance matrix. In particular, in the first experiment for $\gamma = 1000$, only twelve edges are chosen. Relative to the centralized network with the

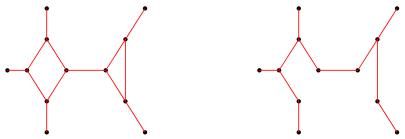


(a) Original network with $n = 12$ nodes (b) $\gamma = 0.1$



(c) $\gamma = 1000$ (d) Distinct edges

Fig. 1: The problem of identification of sparse networks using sample covariance matrix for a network with $n = 12$ nodes.



(a) $\gamma = 0.1$ (b) $\gamma = 1000$

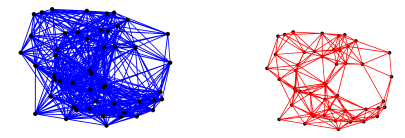
Fig. 2: The problem of identification of sparse networks using sample covariance matrix for a network with $n = 12$ nodes.

vector of the edge weights x_c, the identified sparse network in this case uses only 18.182% of the edges, i.e., $\mathbf{card}(x)/\mathbf{card}(x_c) = 18.182\%$ and achieves a relative error of 53.841%, $(\|X^{-1} - S\|_F)/\|S\|_F = 53.841\%$, with $X = \left(E \operatorname{diag}(x) E^T + (1/n)\,\mathbb{1}\,\mathbb{1}^T\right)$. In the second experiment, the identified sparse network has 16.666% of the potential edges and achieves a relative error of 51.067%.

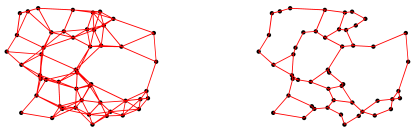
B. Network sparsification

We next use (NI) to find a sparse representation of a dense consensus network. Inspired by [39], we generate a network by randomly distributing nodes in a 10×10 box. A pair of nodes can communicate if the Euclidean distance between them, $d(i, j)$, is not greater than 5 units and the edge connecting them has weight $\exp(-d(i, j))$. The incidence matrix of the identified graph is selected to be equal to the incidence matrix of the given graph; i.e., the sparse network's edges are a subset of the original network's. Figure 3a shows a graph with 50 nodes. We use the reweighted ℓ_1 regularized Gaussian maximum likelihood estimation framework, for 200 logarithmically-spaced values of $\gamma \in [0.01, 100]$ following by the polishing step. The sparse graph topologies identified for different values of γ are shown in figures 3b, 3c, and 3d. As γ increases, the identified graph becomes sparser.

For $\gamma = 0.01$, 222 edges are chosen to be in the sparse estimated network which is only 34.851% of the



(a) Original network with $n = 50$ nodes (b) $\gamma = 0.01$



(c) $\gamma = 0.102$ (d) $\gamma = 1.138$

Fig. 3: The problem of sparsification of a network with $n = 50$ nodes using sample covariance matrix.

637 potential edges. The network with these selected edges achieves a relative error of 29.965%,

$$\frac{\|X^{-1} - S\|_F}{\|S\|_F} = 29.965\%.$$

For the largest value of the sparsity-promoting parameter, $\gamma = 1.138$, only 64 edges are present (10.047% of the potential edges) in the estimated graph that gets a relative error of 207.493%.

To provide a point of comparison, we compare the performance of our algorithm to a simple truncation scheme. In this scheme, the edge with the smallest weights that does not disconnect the network is iteratively removed until the network has the desired sparsity. After identifying the topology in this way, the polishing step optimizes the edge weights of the selected set of edges.

Figure 4 shows the relative errors of our algorithm (in red dashed lines) and the truncation algorithm (in blue solid lines) on a log scale against the number of removed edges. As the number of edges in the estimated graph decreases, the relative error of both algorithms increases. The relative error of network topologies identified by our algorithm is much smaller than the error of those identified by the truncation algorithm, and thus our customized algorithm outperforms the truncation method. In particular, when 573 edges are removed, the relative errors for our customized algorithm and the truncation algorithm are 2.075 and 8.977, respectively.

V. Concluding Remarks

We have developed a method for identifying the topology of an undirected consensus network using available statistical data. In order to promote network sparsity, we introduce a convex optimization framework aimed at finding the solution to the ℓ_1-regularized maximum likelihood problem. This problem is closely related to the problem of sparse inverse covariance estimation that has received significant attention in the literature. In our setup, additional structure arises from the requirement that data is generated by an undirected consensus network. By exploiting the structure

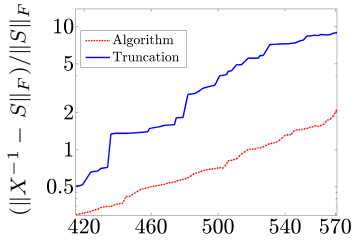


Fig. 4: Relative error in log-scale for the sparsification problem of a network with $n = 50$ nodes using sample covariance matrix.

of the problem, we develop an efficient algorithm based on the sequential quadratic approximation method in which the search direction is determined using coordinate descent with active set strategy. Several examples have been provided to illustrate utility of the method and efficiency of the customized algorithm.

REFERENCES

[1] A. Ghosh and S. Boyd, "Growing well-connected graphs," in *Proceedings of the 45th IEEE Conference on Decision and Control*, 2006, pp. 6605–6611.

[2] R. Dai and M. Mesbahi, "Optimal topology design for dynamic networks," in *Proceedings of the 50th IEEE Conference on Decision and Control*, 2011, pp. 1280–1285.

[3] F. Lin, M. Fardad, and M. R. Jovanovic, "Identification of sparse communication graphs in consensus networks," in *Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing*, 2012.

[4] D. Zelazo, S. Schuler, and F. Allgöwer, "Performance and design of cycles in consensus networks," *Syst. Control Lett.*, vol. 62, no. 1, pp. 85–96, 2013.

[5] M. Fardad, F. Lin, and M. R. Jovanovic, "Design of optimal sparse interconnection graphs for synchronization of oscillator networks," *IEEE Trans. Automat. Control*, vol. 59, no. 9, pp. 2457–2462, September 2014.

[6] T. H. Summers, I. Shames, J. Lygeros, and F. Dörfler, "Topology design for optimal network coherence," in *Proceedings of the 2015 European Control Conference*, 2015, also arXiv:1411.4884.

[7] S. Hassan-Moghaddam and M. R. Jovanovic, "An interior point method for growing connected resistive networks," in *Proceedings of the 2015 American Control Conference*, 2015.

[8] S. Hassan-Moghaddam and M. R. Jovanovic, "Customized algorithms for growing connected resistive networks," in *Proceedings of the 10th IFAC Symposium on Nonlinear Control Systems*, 2016.

[9] P. A. Valdés-Sosa, J. M. Sánchez-Bornot, A. Lage-Castellanos, M. Vega-Hernández, J. Bosch-Bayard, L. Melie-García, and E. Canales-Rodríguez, "Estimating brain functional connectivity with sparse multivariate autoregression," *Phil. Trans. R. Soc. B*, vol. 360, no. 1457, pp. 969–981, 2005.

[10] D. Yu, M. Righero, and L. Kocarev, "Estimating topology of networks," *Physical Review Letters*, vol. 97, no. 18, p. 188701, 2006.

[11] M. M. Zavlanos, A. A. Julius, S. P. Boyd, and G. J. Pappas, "Identification of stable genetic networks using convex programming," in *Proceedings of the 2008 American Control Conference*, pp. 2755–2760.

[12] D. Napoletani and T. D. Sauer, "Reconstructing the topology of sparsely connected dynamical networks," *Physical Review E*, vol. 77, no. 2, p. 026103, 2008.

[13] D. Materassi and G. Innocenti, "Topological identification in networks of dynamical systems," *IEEE Trans. Automat. Control*, vol. 55, no. 8, pp. 1860–1871, 2010.

[14] M. Ayazoglu, M. Sznaier, and N. Ozay, "Blind identification of sparse dynamic networks and applications," in *Proceedings of the 50th IEEE Conference on Decision and Control*, 2011, pp. 2944–2950.

[15] M. Nabi-Abdolyousefi, M. Fazel, and M. Mesbahi, "A graph realization approach to network identification," in *Proceedings of the 51st IEEE Conference on Decision and Control*, 2012, pp. 4642–4647.

[16] D. Materassi and M. V. Salapaka, "On the problem of reconstructing an unknown topology via locality properties of the wiener filter," *IEEE Trans. Automat. Control*, vol. 57, no. 7, pp. 1765–1777, 2012.

[17] N. K. Dhingra, F. Lin, M. Fardad, and M. R. Jovanovic, "On identifying sparse representations of consensus networks," in *Preprints of the 3rd IFAC Workshop on Distributed Estimation and Control in Networked Systems*, Santa Barbara, CA, 2012, pp. 305–310.

[18] M. Siami and N. Motee, "Network sparsification with guaranteed systemic performance measures," in *Preprints of the 5th IFAC Workshop on Distributed Estimation and Control in Networked Systems*, vol. 48, no. 22, 2015, pp. 246–251.

[19] G. F. Young, L. Scardovi, A. Cavagna, I. Giardina, and N. E. Leonard, "Starling flock networks manage uncertainty in consensus at low cost," *PLoS Comput Biol*, vol. 9, no. 1, p. e1002894, 2013.

[20] B. Bamieh, M. R. Jovanovic, P. Mitra, and S. Patterson, "Coherence in large-scale networks: dimension dependent limitations of local feedback," *IEEE Trans. Automat. Control*, September 2012.

[21] K. Scheinberg, S. Ma, and D. Goldfarb, "Sparse inverse covariance selection via alternating linearization methods," in *NIPS*, 2010, pp. 2101–2109.

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

[23] O. Banerjee, L. El Ghaoui, and A. d'Aspremont, "Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data," *J. Mach. Learn. Res.*, vol. 9, pp. 485–516, 2008.

[24] M. Yuan and Y. Lin, "Model selection and estimation in the gaussian graphical model," *Biometrika*, vol. 94, no. 1, pp. 19–35, 2007.

[25] J. Duchi, S. Gould, and D. Koller, "Projected subgradient methods for learning sparse gaussians," *arXiv preprint arXiv:1206.3249*, 2012.

[26] C.-J. Hsieh, I. S. Dhillon, P. K. Ravikumar, and M. A. Sustik, "Sparse inverse covariance matrix estimation using quadratic approximation," in *NIPS*, 2011, pp. 2330–2338.

[27] C.-J. Hsieh, M. A. Sustik, I. S. Dhillon, and P. Ravikumar, "QUIC: Quadratic approximation for sparse inverse covariance estimation," *J. Mach. Learn. Res.*, vol. 15, pp. 2911–2947, 2014.

[28] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *J. Optim. Theory Appl.*, vol. 109, no. 3, pp. 475–494, 2001.

[29] S. Shalev-Shwartz and A. Tewari, "Stochastic methods for ℓ_1 regularized loss minimization," *J. Mach. Learn. Res.*, vol. 12, pp. 1865–1892, June 2011.

[30] A. Saha and A. Tewari, "On the non-asymptotic convergence of cyclic coordinate descent methods," *SIAM Journal on Optimization*, vol. 23, no. 1, pp. 576–601, 2013.

[31] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl*, vol. 14, pp. 877–905, 2008.

[32] M. Mesbahi and M. Egerstedt, *Graph Theoretic Methods in Multiagent Networks*. Princeton University Press, 2010.

[33] B. Rolfs, B. Rajaratnam, D. Guillot, I. Wong, and A. Maleki, "Iterative thresholding algorithm for sparse inverse covariance estimation," in *NIPS*, 2012, pp. 1574–1582.

[34] F. Oztoprak, J. Nocedal, S. Rennie, and P. A. Olsen, "Newton-like methods for sparse inverse covariance estimation," in *NIPS*, 2012, pp. 755–763.

[35] O. Banerjee, L. E. Ghaoui, A. d'Aspremont, and G. Natsoulis, "Convex optimization techniques for fitting sparse gaussian graphical models," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 89–96.

[36] A. d'Aspremont, O. Banerjee, and L. El Ghaoui, "First-order methods for sparse covariance selection," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 1, pp. 56–66, 2008.

[37] C.-J. Hsieh, M. A. Sustik, I. S. Dhillon, P. K. Ravikumar, and R. Poldrack, "Big & quic: Sparse inverse covariance estimation for a million variables," in *NIPS*, 2013, pp. 3165–3173.

[38] S. Hassan-Moghaddam and M. R. Jovanovic, "Topology identification and optimal design of noisy consensus networks," *IEEE Trans. Control Netw. Syst.*, 2016, submitted; also arXiv:1506.03437v2.

[39] N. Motee and A. Jadbabaie, "Optimal control of spatially distributed systems," *IEEE Trans. Automat. Control*, vol. 53, no. 7, pp. 1616–1629, 2008.