Semi-Blind Inference of Topologies and Dynamical Processes over Dynamic Graphs

Vassilis N. Ioannidis, Yanning Shen, and Georgios B. Giannakis
ECE Dept. and Digital Tech. Center, Univ. of Minnesota, Mpls, MN 55455, USA
E-mails: {ioann006, shenx513, georgios} @umn.edu

Abstract—A task of major practical importance in network science is inferring the graph structure from noisy observations at a subset of nodes. Available methods for topology inference typically assume that the process over the network is observed at all nodes. However, application-specific constraints may prevent acquiring network-wide observations. Alleviating the limited flexibility of existing approaches, this work advocates structural models for graph processes and develops novel algorithms for joint inference of the network topology and processes from partial nodal observations. Structural equation models (SEMs) and structural vector autoregressive models (SVARMs) have well-documented merits in identifying even directed topologies of complex graphs; while SEMs capture contemporaneous causal dependencies among nodes, SVARMs further account for time-lagged influences. A batch solver is proposed that iterates between inferring directed graphs that “best” fit the sequence of observations, and estimating the network processes at reduced computational complexity by leveraging tools related to Kalman smoothing. To further accommodate delay-sensitive applications, an online joint inference approach is put forth that even tracks time-evolving topologies. Furthermore, we specify novel conditions for identifying the network topology given partial observations. We prove that the required number of observations for unique identification reduces significantly when the network structure is sparse. Numerical tests with synthetic as well as real datasets corroborate the effectiveness of the proposed approach.

Index Terms—Graph signal reconstruction, topology inference, directed graphs, structural (vector) equation models.

I. INTRODUCTION

Modeling vertex attributes as processes that take values over a graph allows for data processing tasks, such as filtering, inference, and compression, while accounting for information captured by the network topology [25], [41]. However, if the topology is unavailable, inaccurate or even unrelated to the process of interest, performance of the associated task may degrade severely. For example, consider a social graph where the goal is to predict the salaries of all individuals given the salaries of some. Graph-based inference approaches that assume smoothness of the salary over the given graph, may fall short if the salary is dissimilar among friends.

Topology identification is possible when observations at all nodes are available by employing structural models, see e.g., [23]. However, in many real settings one can only afford to collect nodal observations from a subset of nodes due to application-specific restrictions. For example, sampling all nodes may be prohibitive in massive graphs; in social networks individuals may be reluctant to share personal information due to privacy concerns; in sensor networks, devices may report measurements sporadically to save energy; and in gene regulatory networks, gene expression data may contain misses due to experimental errors. In this context, the present paper relies on structural equation models (SEMs) [23], and structural vector autoregressive models (SVARMs) [13] and aims at jointly inferring the network topology and estimating graph signals, given noisy observations at subsets of nodes.

SEMs provide a statistical framework for inferring the structure of a causal network that captures relationships among nodes [18], [23]; see also references therein for related topology identification methods. Linear SEMs have been widely adopted in fields as diverse as sociometrics [19], psychometrics [31], recommender systems [33], and genetics [10]. Conditions for identifying the network topology under the SEM have been also provided [7], [39], but require observations of the process at all nodes. Recently, nonlinear SEMs have been developed to also capture nonlinear interactions [40]. On the other hand, SVARMs postulate that nodes further exert time-lagged dependencies on one another, and are appropriate for modeling multivariate time series [13]. Nonlinear SVARMs have been employed to identify directed dependencies between regions of interest in the brain [38]. Other approaches identify undirected topologies provided that the graph signals are smooth over the graph [15]; or, that the observed process is graph-bandlimited [37]. All these contemporary approaches assume that samples of the graph process are available over all nodes. However, acquiring network-wide observations may incur prohibitive sampling costs, especially for massive networks. On the other hand, network tomography (NT) approaches require partial observations of a specific portion of the network. However, most of NT methods either depend on end-to-end measurements, or assume prior knowledge of the network structure [11], [12], of which neither may be available in general.

Methods for inference of graph signals (or processes), typically assume that the network topology is known and undirected, and the graph signal is smooth, in the sense that neighboring vertices have similar values [42]. Parametric approaches adopt the graph-bandlimited model [6], [32], which postulate that the signal lies in a graph-related B-dimensional subspace; see [29] for time-varying signals. Nonparametric techniques employ kernels on graphs for inference [35], [42]; see also [20] for semi-parametric alternatives. Online data-

The work in this paper was supported by NSF grants 171141, 1500713, and 1442686.
adaptive algorithms for reconstruction of dynamic processes over dynamic graphs have been proposed in [21], where kernel dictionaries are generated from the network topology. However, the aforementioned techniques assume that the topology is known and the graph process is smooth over the graph, which may not hold in practice.

The present paper fills this gap by introducing algorithms based on SEMs and SVARMs for joint inference of network topologies and graph processes over the underlying graph. The approach is semi-blind because it performs the joint estimation task with only partial observations over the network nodes. Instead of reconstructing nodal samples of a graph with given topology, we learn a topology that best fits the observed data, and thus enhance reconstruction performance. Specifically, the contribution is threefold.

C1. A novel approach is proposed for joint inference of directed network topologies and signals over the underlying graph using SEMs. An efficient algorithm is developed with provable convergence at least to a stationary point.

C2. To further accommodate temporal dynamics, we advocate a SVARM to infer dynamic processes and graphs. A batch solver is provided that alternates between topology estimation and signal inference with linear complexity across time. Furthermore, a novel online algorithm is developed that performs real-time joint estimation, and tracks time-evolving topologies.

C3. Analysis of the partially observed noiseless SEM is provided that establishes sufficient conditions for identifiability of the unknown topology. These conditions suggest that the required number of observations for identification reduces significantly when the network exhibits edge sparsity.

The rest of the paper is organized as follows. Sec. II reviews the SEM and SVARM, and states the problem. Sec. III presents a novel estimator for joint inference based on SEMs. Sec. IV develops both batch and online algorithms for inferring dynamic processes and networks using SVARMs. Sec. V presents the identifiability results of the partially observed SEM. Finally, numerical experiments and conclusions are presented in Secs. VI and VII, respectively.

**Notation:** Scalars are denoted by lowercase, column vectors by bold lowercase, and matrices by bold uppercase letters. Superscripts $^T$ and $^{-1}$ respectively denote transpose and inverse; while $I_N$ stands for the $N \times N$ identity matrix. Moreover, $[A]_{i,j}$ denotes a block entry of appropriate size. Finally, if $A$ is a matrix and $x$ a vector, then $\|x\|_A := x^T A^{-1} x$, $\|x\|_2 := x^T x$, $\|A\|_1$ denotes the $L_1$-norm of the vectorized matrix, and $\|A\|_F$ is the Frobenius norm of $A$.

### II. Structural Models and Problem Formulation

Consider a network with $N$ nodes modeled by the graph $G := (V, A)$, where $V := \{v_1, \ldots, v_N\}$ is the set of vertices and $A$ denotes the $N \times N$ adjacency matrix, whose $(n, n')$-th entry $A_{n,n'}$ represents the weight of the directed edge from $v_{n'}$ to $v_n$. A real-valued process (or signal) on $G$ is a map $y_t : V \to \mathbb{R}$. In social networks (e.g., Twitter) over which information diffuses $y_{n,t}$ could represent the popularity (e.g., number of retweets) of a viral story from subscriber $n$ at time $t$. Since real-world networks often exhibit edge sparsity, $A$ has only a few nonzero entries.

#### A. Structural models

The linear SEM [19] postulates that $y_{n,t}$ depends linearly on $\{y_{n',t}\}_{n' \neq n}$, that amounts to

$$y_{n,t} = \sum_{n' \neq n} A_{n,n'} y_{n',t} + \eta_{n,t}$$

(1)

where the unknown $A_{n,n'}$ captures the causal influence of node $v_{n'}$ upon node $v_n$, and $\eta_{n,t}$ accounts for unmodeled dynamics. Clearly, (1) suggests that $y_{n,t}$ is influenced directly by nodes in its neighborhood $N_n := \{v_{n'} : A_{n,n'} \neq 0\}$. With the $N \times 1$ vectors $y_t := [y_{1,t}, \ldots, y_{N,t}]^T$, and $\eta_t := [\eta_{1,t}, \ldots, \eta_{N,t}]^T$, (1) can be written in matrix-vector form as

$$y_t = Ay_t + \eta_t, \quad t = 1, \ldots, T. \quad (2)$$

SVARMs have been successful in a host of applications, including gene regulatory networks [10], and recommender systems [33]. Therefore, the index $t$ does not necessarily indicate time, but may represent different individuals (gene regulatory networks), or movies (recommender systems). An interesting consequence emerges if one considers $\eta_t$ as a random process with $\eta_t \sim \mathcal{N}(0, \sigma^2(t) I_N)$. Since (2) can be written as $y_t = (I_N - A)^{-1} \eta_t = \sum_{n=0}^{\infty} A^n \eta_t$, then $y_t \sim \mathcal{N}(0, \sigma^2(t) \sum_n A^n)^2$ having covariance matrix $C_{y_t} := \sigma^2(t) \sum_n A^n$. Matrices $C_{y_t}$ and $A$ are simultaneously diagonalizable, and hence $y_t$ is a graph stationary process [30].

In order to unveil the hidden causal network topology, SVARMs postulate that each $y_{n,t}$ can be represented as a linear combination of instantaneous measurements at other nodes $\{y_{n',t}\}_{n' \neq n}$, and their time-lagged versions $\{y_{n',t-\ell}\}_{\ell=1}^N$ [13]. Specifically, the following instantaneous plus time-lagged model is advocated

$$y_{n,t} = \sum_{n' \neq n} A_{n,n'}^{(0)} y_{n',t} + \sum_{\ell=1}^N A_{n,n'}^{(1)} y_{n',(t-\ell)} + \eta_{n,t}$$

(3)

where $A_{n,n'}^{(0)}$ captures the instantaneous causal influence of node $n$ upon node $n'$, $A_{n,n'}^{(1)}$ encodes the time-lagged causal influence between them, and $\eta_{n,t}$ accounts for unmodeled dynamics. By defining $y_t := [y_{1,t}, \ldots, y_{N,t}]^T$, $\eta_t := [\eta_{1,t}, \ldots, \eta_{N,t}]^T$, and the $N \times N$ matrices $A^{(0)}$, and $A^{(1)}$
with entries $A^{(0)}_{n,n'}$, and $A^{(1)}_{n,n}$ respectively, the matrix-vector form of (3) becomes

$$y_t = A^{(0)}y_{t-1} + A^{(1)}y_{t-1} + \eta_t, \ t = 1, \ldots, T$$  \hspace{1cm} (4)$$

with $\eta_t = \mu_0 + \mu_0$, and $\mu_0$ considered known. The SVARM in (4) is a better fit for time-series over graphs compared to the SEM in (2), because it further accounts for temporal dynamics of $y_t$ through the time-lagged influence term $A^{(1)}y_{t-1}$. For this reason, SVARMs will be employed for dynamic setups, such as modeling ECoG time series in brain networks, and predicting Internet router delays. The SVARM is depicted in Fig. 1.

B. Problem statement

Application-specific constraints allow only for a limited number of samples across nodes per slot $t$. Suppose that $M_t$ noisy samples of the $t$-th observation vector

$$z_{m,t} = y_{n,m,t} + \epsilon_{m,t}, \ m = 1, \ldots, M_t$$  \hspace{1cm} (5)$$

are available, where $M_t := \{n_1, \ldots, n_{M_t}\}$ contains the indices $1 \leq n_1 \leq \cdots \leq n_{M_t} \leq N$ of the sampled vertices, and $\epsilon_{m,t}$ models the observation error. With $z_t := [z_{1,t}, \ldots, z_{M_t,t}]^T$ and $\epsilon_t := [\epsilon_{1,t}, \ldots, \epsilon_{M_t,t}]^T$, the limited observation model is

$$z_t = M_ty_t + \epsilon_t, \ t = 1, \ldots, T$$  \hspace{1cm} (6)$$

where $M_t$ is an $M_t \times N$ matrix with entries $\{(m, n_m)\}_{m=1}^{M_t}$ set to one, and the rest set to zero.

The broad goal of this paper is the joint inference of the hidden network topology and signals over graphs (JISG) from observations of the latter at subsets of nodes. Given the partial observations $\{z_{i,t}\}_{t=1}^T$ collected in accordance to the sampling matrices $\{M_i\}_{t=1}^T$, one aims at finding the underlying topologies, $A$ for the SEM, or $A^{(0)}$ and $A^{(1)}$ for the SVARM, as well as reconstructing the graph process at all nodes $\{y_{i,t}\}_{t=1}^T$. The complexity of the estimators should preferably scale linearly in $T$. As estimating the topology and $\{y_{i,t}\}_{t=1}^T$ relies on partial observations, this is a semi-blind inference task.

III. JOINTLY INFERRING TOPOLOGY AND SIGNALS

Given $\{z_{i,t}\}_{t=1}^T$ in (6), this section develops a novel approach to infer $A$, and $\{y_{i,t}\}_{t=1}^T$. To this end, we advocate the following regularized least-squares (LS) optimization problem

$$\min_{A \in A, \{y_{i,t}\}_{t=1}^T} \mathcal{F}(A, \{y_{i,t}\}_{t=1}^T) := \sum_{t=1}^T \|y_t - A y_t\|_2^2 + \sum_{t=1}^T \beta \|z_t - M_t y_t\|_2^2 + \rho_e(A)$$  \hspace{1cm} (7)$$

where

$$\rho_e(A) := 2\lambda_1 \|A\|_1 + 2\lambda_2 \|A\|_F^2$$  \hspace{1cm} (8)$$

$\mu \geq 0$ tunes the relative importance of the fitting term; $\lambda_1 \geq 0$, $\lambda_2 \geq 0$ control the effect of the $\ell_1$-norm and the Frobenius-norm, respectively, and $A := \{A : A \in \mathbb{R}^{N \times N}, \{A_{n,n} = 0\}_{n=1}^N\}$. The weighted sum of $\|\cdot\|_1$ and $\|\cdot\|_2$ is the so-called elastic net penalty, which promotes connections between highly correlated nodal measurements. The elastic net targets the “sweet spot” between the $\ell_1$ regularizer that effects sparsity, and the $\|\cdot\|_2$ regularizer, which advocates fully connected networks [44].

Even though (7) is nonconvex in both $A$ and $y_t$ due to the bilinear product $Ay_t$, it is convex with respect to (w.r.t.) each block variable separately. This motivates an iterative block coordinate descent (BCD) algorithm that alternates between estimating $\{y_{i,t}\}_{t=1}^T$ and $A$.

Given $A$, the estimates $\{\hat{y}_{i,t}\}_{t=1}^T$ are found by solving the quadratic problem

$$\min_{\{y_{i,t}\}_{t=1}^T} \sum_{t=1}^T \|y_t - A y_t\|_2^2 + \sum_{t=1}^T \beta \|z_t - M_t y_t\|_2^2$$  \hspace{1cm} (9)$$

where the regularization terms in (7) do not appear. Clearly, (8) conveniently decouples across $t$ as

$$\min_{y_t} g(y_t) := \|\{I_N - A\}y_t\|_2^2 + \|z_t - M_t y_t\|_2^2.$$  \hspace{1cm} (10)$$

The first quadratic in (9) can be written as $\|\{I_N - A\}y_t\|_2^2 = \sum_{n=1}^N (y_{n,t} - \sum_{m \in N_n} A_{n,m} y_{m,t})^2$, and it can be viewed as a regularizer for $y_t$, promoting graph signals with similar values at neighboring nodes. Notice that (9) may not be strongly convex, since $I_N - A$ could be rank deficient. Nonetheless, since $g(y)$ is smooth, (9) can be readily solved via gradient descent (GD) iterations

$$y_t^{(r)} = y_t^{(r-1)} - \frac{\partial}{\partial y_t} g(y_t^{(r-1)}).$$  \hspace{1cm} (11)$$

where $\frac{\partial}{\partial y_t} g(y_t) := \sum_{n=1}^N (I_N - A)^T(I_N - A) + M_t^T M_t y_t - M_t^T z_t$, and $\theta > 0$ is the stepsiz chosen e.g. by the Armijo rule [9]. The computational cost of (10) is dominated by the matrix-vector multiplication of $I_N - A$ with $y_t$, which is proportional to $O(k_{nvs})$, where $k_{nvs}$ denotes the number of non-zero entries of $A$. Moreover, the learned $A$ is expected to be sparse due to the $\ell_1$ regularizer in (7), which renders first-order iterations (10) computationally attractive, especially when graphs are large. The GD iterations (10) are run in parallel across $t$ until convergence to a minimizer of (9).

On the other hand, with $\{\hat{y}_{i,t}\}_{t=1}^T$ available, $A$ is found via

$$\min_{A \in A} \sum_{t=1}^T \|\hat{y}_t - A \hat{y}_t\|_2^2 + \lambda_1 \|A\|_1 + \lambda_2 \|A\|_F^2$$  \hspace{1cm} (12)$$

where the LS observation error in (7) has been omitted. Note that (11) is strongly convex, and as such it admits a unique minimizer. Hence, we adopt the alternating methods of multipliers (ADMM), which guarantees convergence to the global minimum in a finite number of iterations; see e.g. [17].

The derivation of the algorithm is omitted due to lack of space; instead the detailed derivation of a similar ADMM solver for a more general setting will be presented in Sec. IV-A.

The BCD solver for JISG is summarized as Algorithm 1. JISG converges at least to a stationary point of (7), as asserted by the ensuing corollary.
Algorithm 1: Joint Infer. of Signals and Graphs (JISG)

Input: Observations \( \{ z_t \}_{t=1}^T \); sampling matrices \( \{ M_t \}_{t=1}^T \); and regularization parameters \( \{ \lambda_1, \lambda_2 \} \)

1. Initialize: \( \hat{y}_t(0) = M_t^T z_t, \ t = 1, \ldots, T \)
2. while iterates not converge do
3. \( \hat{y}_t(i) = \hat{A}(i) \) from (11) using ADMM.
4. Update \( \{ y_t(i) \}_{t=1}^T \) using (9) and (10).
5. \( i = i + 1 \)
6. end while

Output: \( \{ \hat{y}_t \}_{t=1}^T, \hat{A} \).

Corollary 1. The sequence of iterates \( \{ \{ \hat{y}_t(i) \}_{t=1}^T, \hat{A}(i) \}_{i = 1}^\infty \) resulting from obtaining the global minimizers of (8) and (11), is bounded and converges monotonically to a stationary point of (7).

Proof. The basic convergence results of BCD have been established in [43]. First, notice that all the terms in (7) are differentiable over their open domain except the non-differentiable \( \ell_1 \) norm, which is however separable. These observations establish, based on [43, Lemma 3.1], that \( F(\hat{A}, \{ y_t \}_{t=1}^T) \) is regular at each coordinatewise minimum point \( \hat{A}^*, \{ y_t^* \}_{t=1}^T \), and therefore every such a point is a stationary point of (7). Moreover, \( F(\hat{A}, \{ y_t \}_{t=1}^T) \) is continuous and convex per variable. Hence, by appealing to [43, Theorem 5.1], the sequence of iterates generated by JISG converges monotonically to a coordinatewise minimum point of \( F \), and consequently to a stationary point of (7).

A few remarks are now in order.

Remark 1. A popular alternative to the elastic net regularizer is the nuclear norm \( \rho(A) = \| A \|_n \), that promotes low rank of the learned adjacency matrix - a well-motivated attribute when the graph is expected to exhibit clustered structure [14].

Remark 2. Oftentimes, prior information about \( G \) may be available, e.g. the support of \( A \); nonnegative edge weights \( A_{n,n'} \geq 0, \forall n, n' \); or, the value of \( A_{n,n'} \) for a subset of nodes \( n, n' \in V \). In a social network for example, one may reveal only a subset of his contacts and conceal another due to privacy concerns. Such prior information can be easily incorporated in (7) by adjusting \( A \), and the ADMM solver accordingly.

Remark 3. The estimator in (8) that relies on SEMs is capable of estimating functions over directed graphs as well as undirected ones, while kernel-based approaches [42] and estimators that rely on the graph Fourier transform [41] are usually confined to undirected graphs.

Remark 4. In real-world networks, sets of nodes may depend upon each other via multiple types of relationships, which ordinary networks cannot capture [24]. Consequently, generalizing the traditional single-layer to multilayer networks that organize the nodes into different groups, called layers, is well motivated. Different layers could also capture different forms of social interactions, such as friendship-, family-, or coworker-ties. Such layer structure can be incorporated in (7) via appropriate regularization. One may effect different sparsity levels for networks at different layers or give appropriate weight to each graph layer regularizer for estimating the network process; see e.g. [22]. Thus, the JISG estimator can also accommodate multilayer graphs.

IV. JOINTLY INFER GRAPHS AND PROCESSES OVER TIME

SEMs capture contemporaneous causal dependencies among nodes at the same time instant cf. (2). However, real-world networks often involve processes that vary over time, with dynamics not captured by SEMs. This section considers an alternative based on SVARMs that further models time-lagged influences among nodes (4) and allows for joint inference of dynamic network processes and graphs.

A. Batch Solver for JISG over time

Given \( \{ z_t, M_t \}_{t=1}^T \), this section develops an efficient approach to infer \( \{ A(0), A(1) \} \), and \( \{ y_t \}_{t=0}^T \). Clearly, to cope with the undetermined system of equations (4) and (6), one has to exploit the structure in \( A(0) \) and \( A(1) \). This prompts the following regularized LS objective

\[
\min_{A(0) \in A, A(1) \in A} \sum_{t=0}^T \| y_t - A(0)y_t - A(1)y_{t-1} \|_2^2 + \rho_c(A(0)) + \rho_e(A(1))
\]

\[
+ \| y_0 - \mu_0 \|_2^2 + \sum_{t=1}^T \frac{\mu}{M_t} \| z_t - M_t y_t \|_2^2
\]

where \( \mu > 0 \) is a regularization scalar weighting the fit to the observations, and \( \rho_c(A) \) is the elastic net regularizer for the connectivity matrices. The first sum accounts for the LS fitting error of the SVARM, and the second LS cost accounts for the initial conditions. The third term sums the measurement error over \( t \). Finally, the elastic net penalty terms \( \rho_c(A(0)) \), and \( \rho_e(A(0)) \) favor connections among highly correlated nodes; see also discussion after (7).

The optimization problem in (12) is nonconvex due to the bilinear terms \( A(0)y_t \) and \( A(1)y_{t-1} \); nevertheless, it is convex w.r.t. each of the variables separately. Next, an efficient algorithm based on BCD is put forth that provably attains a stationary point of (12). With \( A(0) \), and \( A(1) \) available, the following objective yields estimates

\[
\{ \hat{y}_t \}_{t=0}^T := \arg \min_{\{ y_t \}_{t=0}^T} \sum_{t=1}^T \| y_t - A(0)y_t - A(1)y_{t-1} \|_2^2 + \sum_{t=1}^T \frac{\mu}{M_t} \| z_t - M_t y_t \|_2^2
\]

\[
+ \| y_0 - \mu_0 \|_2^2 + \sum_{t=1}^T \frac{\mu}{M_t} \| z_t - M_t y_t \|_2^2
\]

where \( \hat{y}_t \) denotes the estimate of \( y_t \) given \( \{ z_t \}_{t=1}^T \). Different from (8), the time-lagged dependencies \( A(1)^{-1}y_{t-1} \) couple the objective in (13) across \( t \). Upon defining \( \hat{A} := \hat{A}(0)^{-1} \) that is assumed invertible, \( F := \hat{A}(0)^{-1} \), and \( Q := (\hat{A}(1)\hat{A})^{-1} \), we can express (13) equivalently as

\[
\{ \hat{y}_t \}_{t=0}^T := \arg \min_{\{ y_t \}_{t=0}^T} \sum_{t=1}^T \| y_t - M_t y_{t-1} \|_2^2 + \sum_{t=1}^T \frac{\mu}{M_t} \| z_t - M_t y_t \|_2^2
\]

\[
+ \| y_0 - \mu_0 \|_2^2 + \sum_{t=1}^T \frac{\mu}{M_t} \| z_t - M_t y_t \|_2^2 + \frac{1}{M_0} \| y_0 - \mu_0 \|_2^2
\]
Proposition 1. The minimizers of (13) $\{\hat{y}_t\}_{t=0}^T$ are given iteratively by the RTS smoother summarized as Algorithm 2.

Algorithm 2: RTS smoother

Input: $F, \Omega, \psi, \gamma_0[0]$; $\Sigma[0][0] \cdot \{z_t\}_{t=1}^T$; $\{M_t\}_{t=1}^T$

for $t = 1, 2, \ldots, T$ do
KF1. $\hat{y}_{t|t-1} = F \hat{y}_{t-1|t-1}$
KF2. $\Sigma_{t|t-1} = F \Sigma_{t-1|t-1} F^T + Q$
KF3. $G_t^2 = \Sigma_{t|t-1} M_t^1 ((M_t^1 / \mu)M_t + M_t \Sigma_{t-1|t-1} M_t^T)^{-1}$
KF4. $\Sigma_t = (I - G_t^2 M_t) \Sigma_{t-1|t-1}$
KF5. $\hat{y}_t = \hat{y}_{t|t-1} + G_t^2 z_t$
end for

Output: $\{\hat{y}_t\}_{t=0}^T$

\[ + \|y_0 - \mu_0\|_2^2 + \sum_{t=1}^T \frac{\mu}{M_t} z_t - M_t y_t \|_2^2 \] (14)

The minimizer of (14) can be attained by

\[ \hat{\psi} = \arg \min_{\psi} \|y_0 - \mu_0\|_2^2 + \sum_{t=1}^T \frac{\mu}{M_t} z_t - M_t y_t \|_2^2 \]

with

\[ \Omega = \text{bdiag}\{I_1, Q_1, \ldots, Q_T, M_1 / \mu I_{M_1}, \ldots, M_T / \mu I_{M_T}\} \]

and the backward direction is performed by Kalman smoothing (steps KS1-KS2) [34]. The algorithm smoothly the state estimates over the interval $[1, T]$. Each step incurs complexity at most $O(N^3)$, and hence the overall complexity for estimating $\{\hat{y}_t\}_{t=0}^T$ is $O(N^3 T^3)$, which scales favorably for large $T$. For solving (13), one should initialize the RTS by $\hat{y}_0[0] = \mu_0$, and $\Sigma[0][0] = I_N$.

To estimate the adjacency matrices given $\{\hat{y}_t\}_{t=0}^T$, consider the following problem

\[ \min_{A(0) \in A(1)} \sum_{t=1}^T \|y_t - A(0) \hat{y}_t - A(1) \hat{y}_{t-1}\|_2^2 + \rho_e A(0)^2 + \rho_e A(1)^2 \] (16)

The objective in (16) is convex albeit nonsmooth, and hence ADMM can be adopted to obtain $\hat{A}(0)$ and $\hat{A}(1)$. The ADMM solver is summarized as Algorithm 3, and its derivation is deferred to the Appendix.

The overall procedure for joint inference of signals and graphs over time (JISGoT) is tabulated as Algorithm 4. Convergence of JISGoT is asserted in the following corollary, the proof of which is similar to Corollary 1, and hence is omitted.

Corollary 2. The sequence of iterates $\{\{\hat{y}_t(i)\}_{t=0}^T, A(0)(i), A(1)(i)\}_{i}$, resulting from obtaining the global minimizers of (13) and (16), is bounded and converges monotonically to a stationary point of (12).

Note that (12) corresponds to a batch problem formulation, i.e. all observation must be available $\{z_t\}_{t=1}^T$ before estimating $\{\hat{y}_t\}_{t=1}^T$ and $\hat{A}(0), \hat{A}(1)$. Next, we introduce an updated formulation that allows for online inference.

B. Fixed-lag solver for online JISGoT

JISGoT performs fixed-interval smoothing, since the whole batch $\{z_t\}_{t=1}^T$ has to be available to learn $\{\hat{y}_t\}_{t=1}^T$. Albeit useful for applications such as processing electroencephalograms, and analysis of historical trade transaction data, this batch solver is not suitable for online applications, such as stock market prediction, analysis of online social networks, and propagation of cascades over interdependent power networks. Such applications enforce strict delay constraints and require estimates within a window or fixed lag [5].
Driven by the aforementioned delay constraints, the goal here is to estimate $\{y_t\}_{t=t+w}$, $A(0)$, $A(1)$, relying upon observations up to time $t+w$, with $w$ denoting the affordable delay window length. Supposing that a KF (cf. Algorithm 2) has been run up to time $t$ to yield estimates $\hat{y}_{t|t}$ and $\Sigma_{t|t}$, the desired estimates $\hat{A}(0), \hat{A}(1)$ can be obtained at $t+w$ by solving the following problem

$$
\arg\min_{\{\hat{A}(0), \hat{A}(1)\}\in\mathcal{A}, t\rightarrow t+1} \sum_{t'=t}^{t+w} \frac{1}{2} \left( \|y_{t'} - \hat{A}(0)y_{t'} - \hat{A}(1)y_{t'-1}\|_2^2 + \mu_c \|y_{t'} - \hat{y}_{t'|t}\|_2^2 + \mu_A \|\hat{A}(0) - \hat{A}(0)_{t-w}\|_F + \mu_A \|\hat{A}(1) - \hat{A}(1)_{t-w}\|_F \right)
$$

where $\hat{A}(0)_{t-w}$ and $\hat{A}(1)_{t-w}$ are the solutions to (17) at $t-w-1$, and $\mu_A$ controls the effect of the LS terms that promote slow-varying $\{\hat{A}(0)_{t}\}_{t=0,1}$. Similar to (12), the fixed-lag objective (17) will be solved via a BCD algorithm to a stationary point. Observe that solving (17) for $\hat{y}_{t|w}$ is a special case of the fixed-interval objective in (13), when the initial condition on the state, namely $\hat{y}_{t|t}$ and $\Sigma_{t|t}$, are given by the RTS algorithm, and the state is smoothed over the interval $[t, t+w]$; see [5] for details on the fixed-lag smoother. Solving for $\{\hat{A}(0)_{t}\}_{t=0,1}$ entails the additional quadratic terms $\|\hat{A}(0) - \hat{A}(0)_{t-w}\|_F$ relative to (16) that can be easily incorporated in Algorithm 3.

Thus, one can employ Algorithm 4 with minor modifications to solve the fixed-lag objective (17) and estimate $\{\hat{A}(0), \hat{A}(1)\}_{t\rightarrow t+w}$. As a convenient byproduct, the novel online estimator (17) tracks dynamic topologies from the time-varying estimates $\hat{A}(0)$ and $\hat{A}(1)$. This is well-motivated when the process is non-stationary, and the underlying topologies change over time.

Remark 5. Although this section builds upon the SVARM in (4) that accounts only for a single time-lag, the proposed algorithms can be readily extended to accommodate SVARMs with multiple time-lags, i.e., $y_t = \sum_{\xi=0}^{\Xi} A(\xi)y_{t-\xi} + \eta_t$. By defining $\Xi \times 1$ extended vector process $\tilde{y}_t = [y_t^T, y_{t-1}^T, \ldots, y_{t-\Xi+1}^T]^T$, the $\Xi \times \Xi$ block matrix $A(1)$, with entries $A(1)_{\xi,\xi} = A(\xi), \xi = 1, \ldots, \Xi$, and zero otherwise, the $\Xi \times \Xi$ block matrix $A(0)$, with entries $A(0)_{\xi,\xi} = A(0), \xi = 1, \ldots, \Xi$, and zero otherwise, $A_{n,n} = I_N$ for $n = 2, \ldots, \Xi$ and zero otherwise, and the $\Xi N \times 1$ error vector $\eta_t = [\eta_{1}^T, \ldots, \eta_{\Xi}^T]^T$, the general SVARM can be written as $\tilde{y}_t = \hat{A}(0)\tilde{y}_t + \hat{A}(1)\tilde{y}_{t-1} + \hat{\eta}_t$, which resembles a single time-lag SVARM in (4).

V. IDENTIFIABILITY ANALYSIS

SEMs are linear systems of equations and identification of the topology (parameter $A$) relates to system identification (SI) [27]. Identifiability of SEMs however, pertains to deterministic linear models (see e.g. [7]), and does not require knowledge of the underlying probability distribution function of the data. Some early results in identifying the parameters of an SEM with exogenous information were reported in [7]. Although interesting, the analysis in [7] considers that the process is observable at all nodes, and hence cannot be applied here. This section provides novel results on the identifiability of the network topology given observations at subsets of nodes under the noise-free SEM.

Specifically, in the absence of noise ($\eta_t = \epsilon_t = 0$) (2) and (6) can be written as

$$
y_t = A\tilde{y}_t, \quad \tilde{y}_t = M_t\tilde{y}_t, \quad t = 1, \ldots, T
$$

where $M_t = \{0, 1\}^N \times N$ with $[M_{t}]_{n,n} := M_{t_{n,n}} = 1$ if $n \in M_t$, and zero otherwise, and $\tilde{y}_t \in \mathbb{R}^N$ with $\tilde{y}_n = y_n$ if $n \in M_t$, and zero if $n$ is not sampled at $t$. The $N \times T$ matrix $\tilde{Z} := \{\tilde{z}_t, \tilde{Z}_T\}$ collects all the observations.

Theorem 1. Under as1 and as2, the adjacency matrix $A$ can be uniquely identified from (18a) and (18b).

To establish identifiability results, we rely on a couple of assumptions.

as1. Matrix $A$ has at most $S$ non-zero entries per row.

as2. For any subset of row indices $R := \{n_1, n_2, \ldots, n_{2S}\}$ with $1 \leq n_i \leq N$, there exists a subset of column indices $C := \{t_1, t_2, \ldots, t_{2S}\}$ with $1 \leq t_i \leq T$ that forms the $2S \times 2S$ matrix $\tilde{Z}_{CR}$ with entries $[\tilde{Z}_{CR}]_{i,j} = \tilde{z}_{n_i,t_j}$ that is fully observable; meaning $\tilde{z}_{n_i,t_j} = y_{n_i,t_j}$ and $[M_{t}]_{n_i,n_i} = \tilde{M}_{n_i,n_i} = 1, \forall n_i \in R, t_j \in C$, and satisfies $\text{Kruskal}(\tilde{Z}_{CR}) = 2S$.

Proof. Combining (18a) and (18b) yields

$$
\tilde{z}_t = M_t A\tilde{y}_t, \quad t = 1, \ldots, T.
$$

Collecting the equations over $t$ leads to the $N \times T$ matrix $\tilde{Z} = [M_1 A\tilde{y}_1, \ldots, M_T A\tilde{y}_T]$. Considering the $n$-th row of $\tilde{Z} := [\tilde{z}_1, \ldots, \tilde{z}_N]^T$ one obtains

$$
\tilde{z}_n = [\tilde{M}_{n,1} a_{0,n}, \tilde{M}_{n,2} a_{1,n}, \ldots, \tilde{M}_{n,T} a_{T,n}]^T = (Y_n^\top)^T a_n
$$

where $Y_n^\top = [\tilde{M}_{n,1} a_{0,n}, \ldots, \tilde{M}_{n,T} a_{T,n}]$ and $a_n^\top$ denotes the $n$-th row of $A$. To argue by contradiction, suppose there exists a vector $a_n \in \mathbb{R}^{N \times 1}$ with $S$ nonzero entries, and $a_n \neq a_n^\top$, such that the following holds

$$
\tilde{z}_n = (Y_n^\top)^T a_n.
$$
Combining (21) with (20) one arrives at
\[
0_T = (Y^R_n)^\top (D_n - a_n).
\] (22)

Since \(a_n\) and \(D_n\), both have \(S\) nonzero entries, \(D_n - a_n\) has at most \(2S\) nonzero entries. Without loss of generality assume that \(R' := \{t'_1, t'_2, \ldots, t'_S\}\) contains all the indices of nonzero entries in \(D_n - a_n\). Let \(c_{R'} := D_n - a_n\), and \(e_{R'} \in \mathbb{R}^{2S 	imes 1}\) denote the sub-vector containing all entries of \(D_n - a_n\) indexed by \(R'\). Hence, (22) can be rewritten as
\[
0_T = ([Y^R_n]_{R'})^\top c_{R'}.
\] (23)

where \([Y^R_n]_{R'}\) selects the rows of \(Y^R_n\) indexed by \(R'\). According to as3, for any set of row indices \(R'\) we can find a set of column indices such that \(Z_{C_{R'}}\) is fully observable, and thus \(\hat{Z}_{C_{R'}} = [Y^R_n]_{R'} = [Y]_{C_{R'}}\). As a result, it holds that
\[
0_{2S} = \hat{Z}_{C_{R'}} c_{R'}.
\] (24)

Since \(\text{Kruskal}(\hat{Z}_{C_{R'}}) = 2S\), any subset of \(2S\) columns of \(\hat{Z}_{C_{R'}}\) is linearly independent, which implies \(c_{R'} = 0\), and by definition \(D_n - a_n\), which leads to a contradiction. The analysis from (20) to (24) holds for all \(n = 1, \ldots, N\). \(\square\)

Remark 6. For time-varying topologies identification could still be achieved in the special case where \(A_t\) is piece-wise constant i.e. \(A_t = A_{\tau_t}, \forall t \in \tau_t\), where \(\tau_t\) is a time interval. Therefore, one can still apply Theorem 1 per \(\tau_t\), that requires at least \(2S\) observations at each node over \(\tau_t\) for identifiability of \(A_{\tau_t}\).

Corollary 3 establishes the sufficient condition for identifying the network structure given full observations for some \(t\).

Remark 7. Some works have analyzed SEMs with missing data [4], [16], [28]. The methods summarized in [4] and [16] either “delete data samples” that contain misses, or, they assume that the support (location of nonzero entries) of \(A\) is known. Meanwhile, [28] provide a sparsity-promoting algorithm for SEMs and vector autoregressive models (VARM) of order one with static parameters (\(A\)), as well as asymptotic performance analysis. The present paper’s contributions are: i) without deleting samples, inference relies on all partial observations; ii) missing process values can be interpolated; iii) topology ID is accomplished without prior knowledge on the support of \(A\); iv) recovery guarantees for topology ID are provided based on finite number of observations; and v) utilizing the Kalman smoother, online scalable algorithms are developed, and can be applied even for higher-order SVARMs.

VI. NUMERICAL TESTS

The tests in this section evaluate the performance of the proposed joint inference approach in comparison with state-of-the-art graph signal inference and topology identification techniques using synthetic and real data.

The network topology performance is measured by the edge identification error rate (EIER), defined as
\[
\text{EIER} := \frac{||S - \hat{S}||_0}{N(N - 1)} \times 100\%.
\]

with the operator || \cdot ||_0 denoting the number of nonzero entries of its argument, and \(S (\hat{S})\) the support of \(A (\hat{A})\). For the estimated adjacency an edge is declared present if \(A_{n,n'}\) exceeds a threshold chosen to yield the smallest EIER. The inference performance of JSG is assessed by comparing with the normalized mean-square error
\[
\text{NMSE} := \frac{1}{T} \sum_{t=1}^{T} \frac{||\hat{y}_t - y_t||_2^2}{||y_t||_2^2}.
\]

Parameters \(\mu, \lambda_1\) and \(\lambda_2\) are selected via cross validation. The software used to conduct all experiments is MATLAB. All results represent averages over 10 independent Monte Carlo runs. Unless otherwise stated, \(M_t\) is chosen uniformly at random without replacement over \(V\) for each \(t\) with constant size over time; that is, \(M_t = M, \forall t\).
A. Numerical tests on synthetic data

First, a synthetic network of size \( N = 81 \) was generated using the Kronecker product model, that effectively captures properties of real graphs [26]. It relies on the “seed matrix”

\[
D_0 := \begin{bmatrix}
0.6 & 0.1 & 0.7 \\
0.3 & 0.1 & 0.5 \\
0 & 1 & 0.1
\end{bmatrix}
\]

that produces the \( N \times N \) matrix as \( D := D_0 \otimes D_0 \otimes D_0 \otimes D_0 \), where \( \otimes \) denotes Kronecker product. The entries of \( A \) were selected as \( A_{n,n'} \sim \text{Bernoulli}(D_{n,n'}) \) \( \forall n, n' \), and the resulting matrix was rendered symmetric by adding its transpose. The graph signals were generated using the graph-bandlimited model \( y_t = \sum_{i=1}^{10} \gamma_i(t) u_i^{(t)} \), \( t = 1, \ldots, T \), where \( T = 100 \), \( \gamma_i(t) \sim N(0, 1) \), and \( \{u_i^{(t)}\}_{i=1}^{10} \) are the eigenvectors associated with the 10 smallest eigenvalues of the Laplacian matrix \( L := \text{diag}\{A1\} - A \). We generate Gaussian noise \( \epsilon_{n,t} \) with signal to noise ratio (SNR)=10dB and add it to the observations according to (5).

The compared estimators for graph signal inference include the bandlimited estimator (BL) [6], [32] with bandwidth \( B \); and the multi-kernel learning (MKL) estimator that employs a dictionary comprising 100 diffusion kernels with parameter \( \sigma^2 \) uniformly spaced between 0.01 and 2, and selects the kernel that “fits” best the observed data [36]. These reconstruction algorithms assume the topology is known and symmetric, which may not always be the case. To capture model mismatch, BL and MKL use \( A + E \) with \( E_{n,n'} \sim N(0, 0.05) \) instead of \( A \). Fig. 3 shows the NMSE of various approaches with increasing \( M \), where \( M_t = M \ \forall t \), and the baseline is the MKL that considers the true topology \( A \). The reconstruction performance of JISG is superior compared to that of BL and MKL, and matches the baseline performance. Moreover, the reported CPU time of JISG at 0.12 seconds was an order of magnitude faster than that of the MKL baseline at 1.6 seconds.

For the same simulation setting, the topology inference performance was evaluated, by comparing with the elastic net (EN) SEM that identifies the network topology from observations across all nodes, meaning \( \{z_t = y_t\}_{t=1}^{T} \). Fig. 4 plots the EIER with increasing \( M \) for JISG while EN-SEM uses \( M = N \). The semi-blind novel approach achieves similar performance with the baseline, which can not cope with missing nodal measurements.

B. Gene regulatory network identification

Further tests were conducted using real gene expression data [10]. Nodes in this network represent \( N = 39 \) immune-related genes, while the measurements consist of gene expression data from \( T = 69 \) unrelated Nigerian individuals. The graph process \( y_{n,t} \) measures the expression level of gene \( n \) for individual \( t \). This experiment evaluates the topology inference performance of JISG with \( M_t = 31 \) genes for all individuals sampled at random. Since no ground-truth topology is available here, the estimated adjacency of EN-SEM, that relies on all the observations, was used for comparison. Fig. 5 depicts heatmaps of the estimated adjacencies. As observed, JISG learns a topology similar to that identified by EN-SEM, and imputes the missing values with NMSE = 0.017. Therefore, our joint inference approach is capable of revealing causal dependencies even when gene expression data contain missing values.
C. Temperature prediction

Consider the National Climatic dataset, which comprises hourly temperature measurements at \( N = 109 \) measuring stations across the continental United States in 2010 [1]. The value \( y_{n,t} \) represents here the \( t \)-th temperature sample recorded at the \( n \)-th station. For evaluating the JISGoT the cumulative NMSE (cNMSE) was used

\[
cNMSE(T) := \frac{\sum_{t=1}^{T} \| y_t - \hat{y}_t \|^2}{\sum_{t=1}^{T} \| y_t \|^2}.
\]

Next, the proposed method is compared to the graph-bandlimited approach [6], [32] for different bandwidth values \( B \), where a time-invariant graph was constructed as in [35], based on geographical distances. Fig. 6 reports the cNMSE performance of the estimators with \( M = 76 \) for increasing \( T \). JISGoT learns the latent topology among sensors and outperforms the band-limited estimator since the temperature may not adhere to the band-limited model for the geographical graph.

Fig. 7 shows the cNMSE of JISGoT with variable \( M \). As expected, the performance improves with increasing number of samples, while with just 30% sampled stations the normalized reconstruction error is only 0.018. Hence, JISGoT can be employed to effectively predict missing sensor measurements.

D. GDP prediction

This experiment is carried over the gross domestic product (GDP) dataset [2], which comprises GDP per capita for \( N = 127 \) countries for the years 1960-2016. The process \( y_{n,t} \) now denotes the GDP reported at the \( n \)-th country and \( t \)-th year for \( t = 1960, \ldots, 2016 \). Fig. 8 shows the cNMSE performance of our joint approach for different \( M \). The semi-blind estimator unveils the latent connections among countries, while it reconstructs the GDP with cNMSE=0.05 when 60% samples are available.

Fig. 9 depicts the true values, along with the GDP estimates of Greece for \( M = 89 \), which corroborates the effectiveness of JISGoT in predicting the GDP evolution and henceforth facilitating economic policy planning.

E. Network delay prediction

The last dataset records measurements of path delays on the Internet2 backbone [3]. The network comprises 9 end-
nodes and 26 directed links. The delays are available for $N = 70$ paths per minute. Function $y_{n,t}$ denotes the delay in milliseconds measured at the $n$-th path and $t$-th minute.

The proposed JISGoT will be evaluated in estimating the delay over the network from $M = 49$ randomly sampled path delays. To that end, delay maps are traditionally employed, which depict the network delay per path over time and enable operators to perform troubleshooting. The paths for the delay maps in Fig. 11 are sorted in increasing order of the true delay at $t = 1$. Clearly, the delay map recovered by JISGoT in Fig. 10b visually resembles the true delay map in Fig. 10a.

VII. CONCLUSIONS AND FUTURE WORK

This paper puts forth a novel framework based on SVARMs and SEMs to jointly infer sparse directed network topologies, and even dynamic graph processes. Efficient minimization approaches are developed with provable convergence that alternate between reconstructing the network processes and inferring the topologies using ADMM. The framework was broadened to facilitate real-time sequential joint-estimation by employing a fixed-lag solver. Recognizing the challenges related to partially observed processes, conditions under which the network can be uniquely identified were derived. Numerical tests on synthetic and real data-sets demonstrate the competitive performance of JISG and JISGoT in both inferring graph signals and the underlying network topologies.

Future research could pursue learning nonlinear models of the network processes, alternative solvers that further account for nonlinear constraints as the fast iterative shrinkage-thresholding algorithm [8], and distributed implementation of JISG, which is well-motivated, especially when dealing with large-scale networks.

APPENDIX

A. ADMM solver for (16)

Towards deriving the ADMM solver, consider $Y_T := [Y_1, Y_2, \ldots, Y_T]^T$, $Y_{T-1} := [Y_0, Y_1, \ldots, Y_{T-1}]^T$, and the auxiliary variables $\Psi^{(0)}$ and $\Psi^{(1)}$. Then, re-write (16) as

$$\min_{A^{(0)}, A^{(1)}, \Psi^{(0)}, \Psi^{(1)}} \frac{1}{2} \|Y_T - A^{(0)}Y_T - A^{(1)}Y_{T-1}\|_F^2 + \lambda_1^{(0)}\|\Psi^{(0)}\|_1 + \lambda_2^{(0)}\|\Psi^{(1)}\|_1$$

s.t. $A^{(0)} = \Psi^{(0)} - \text{diag}(\Psi^{(0)})$, $A^{(1)} = \Psi^{(1)}$. (25)

The augmented Lagrangian of (25) is

$$\mathcal{L} = \frac{1}{2} \|Y_T - A^{(0)}Y_T - A^{(1)}Y_{T-1}\|_F^2 + \lambda_1^{(0)}\|\Psi^{(0)}\|_1 + \lambda_2^{(0)}\|\Psi^{(1)}\|_1$$

$$+ \lambda_1^{(1)}\|\Psi^{(1)}\|_1 + \lambda_2^{(1)}\|\Psi^{(1)}\|_1$$

$$+ \frac{1}{2}\|A^{(0)} - \Psi^{(0)} + \text{diag}(\Psi^{(0)})\|_F^2$$

$$\ \+ \frac{1}{2}\|A^{(1)} - \Psi^{(1)}\|_F^2 \ \(26\)$$

where $U^{(0)}$ and $U^{(1)}$ denote Lagrange multiplier matrices, while $\rho > 0$ is the penalty parameter. Henceforth, square brackets denote ADMM iteration indices. The ADMM update for $A^{(0)}$ results from $\partial \mathcal{L}/\partial A^{(0)} = 0$ that gives

$$A^{(0)}[i][R_T + (\lambda_s^{(0)} + \rho)1_N] = \frac{R_T - A^{(1)}[i - 1]C - U^{(0)}[i - 1] + \rho\Psi^{(0)}[i - 1]}{1 + \rho} \ \(27\)$$

where $R_T := Y_T^T Y_T$, $C := Y_{T-1}^T Y_{T-1}$. Similarly for $A^{(1)}$, taking $\partial \mathcal{L}/\partial A^{(1)} = 0$ results to

$$A^{(1)}[i][R_{T-1} + (\lambda_s^{(1)} + \rho)1_N] = C - A^{(0)}[i]C - U^{(1)}[i - 1] + \rho\Psi^{(1)}[i - 1] \ \(28\)$$

Finally, the Lagrange multiplier updates are given by

$$U^{(0)}[i] = U^{(0)}[i - 1] + \rho(1^{(0)}[i] - \Psi^{(0)}[i]) \ \(31a\)$$

$$U^{(1)}[i] = U^{(1)}[i - 1] + \rho(1^{(1)}[i] - \Psi^{(1)}[i]) \ \(31b\)$$

The complexity of (27) and (28) is $O(N^3)$, while for (29)- (31) is $O(N^2)$ that leads to an overall per ADMM iteration complexity of $O(N^3)$, governed by the updates for $A^{(0)}$ and $A^{(1)}$. That brings the overall complexity of the algorithm to $O(IN^3)$, where $I$ is the number of required ADMM iterations until convergence.

REFERENCES


Fig. 10: True and estimated network delay map for $N = 70$ paths ($\mu = 100$, $\lambda_1 = 0.01$, $\lambda_2 = 0.01$)
V. N. Ioannidis (S’16) received his Diploma in Electrical and Computer Engineering from the National Technical University of Athens, Greece, in 2015, and the M.Sc. degree in Electrical Engineering from the University of Minnesota (UMN), Twin Cities, Minneapolis, MN, USA, in 2017. Currently, he is working towards the Ph.D. degree in the Department of Electrical and Computer Engineering at the University of Minnesota, Twin Cities. From 2014 to 2015, he worked as a software engineer for Oracle in Athens, Greece, and received a Performance Excellence award. His research interests include machine learning, big data analytics, and network science.

Y. Shen received her B.Sc. and M.Sc. degrees in Electrical Engineering from the University of Electronic Science and Technology of China, in 2011 and 2014, respectively. Since September 2014, she has been working towards her Ph.D. degree with the Department of Electrical and Computer Engineering, University of Minnesota (UMN). Her research interests include data science, network science, machine learning, optimization and statistical signal processing. She received the UESTC distinguished B.Sc. thesis Award in 2011, and distinguished M.Sc. Thesis Award in 2014. She was a Best Student Paper Award finalist of the 2017 IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing. She was selected as Rising Stars in EECS by Stanford University in 2017, and received the UMN Doctoral Dissertation Fellowship in 2018.

G. B. Giannakis (Fellow’97) received his Diploma in Electrical Engr. from the Ntl. Tech. Univ. of Athens, Greece, 1981. From 1982 to 1986 he was with the Univ. of Southern California (USC), where he received his MSc. in Electrical Engineering, 1983, MSc. in Mathematics, 1986, and Ph.D. in Electrical Engr., 1986. He was a faculty member with the University of Virginia from 1987 to 1998, and since 1999 he has been a professor with the Univ. of Minnesota, where he holds an ADC Endowed Chair, a University of Minnesota McKnight Presidential Chair in ECE, and serves as director of the Digital Technology Center.

His general interests span the areas of statistical learning, communications, and networking - subjects on which he has published more than 450 journal papers, 750 conference papers, 25 book chapters, two edited books and two research monographs (h-index 140). Current research focuses on Data Science, and Network Science with applications to the Internet of Things, social, brain, and power networks with renewables. He is the (co-) inventor of 32 patents issued, and the (co-) recipient of 9 best journal paper awards from the IEEE Signal Processing (SP) and Communications Societies, including the G. Marconi Prize Paper Award in Wireless Communications. He also received Technical Achievement Awards from the SP Society (2000), from EURASIP (2005), a Young Faculty Teaching Award, the G. W. Taylor Award for Distinguished Research from the University of Minnesota, and the IEEE Fourier Technical Field Award (inaugural recipient in 2015). He is a Fellow of EURASIP, and has served the IEEE in a number of posts, including that of a Distinguished Lecturer for the IEEE-SPS.