This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2020.2982833, IEEE Transactions on Signal Processing

Graph-based Learning under Perturbations via Total Least-Squares

Elena Ceci, *Student Member, IEEE*, Yanning Shen, *Member, IEEE*, Georgios B. Giannakis, *Fellow, IEEE*, Sergio Barbarossa, *Fellow, IEEE*

***Abstract*—Graphs are pervasive in different fields unveiling complex relationships between data. Two major graph-based learning tasks are topology identification and inference of signals over graphs. Among the possible models to explain data interdependencies, structural equation models (SEMs) accommodate a gamut of applications involving topology identification. Obtaining conventional SEMs though requires measurements across nodes. On the other hand, typical signal inference approaches 'blindly trust' a given nominal topology. In practice however, signal or topology perturbations may be present in both tasks, due to model mismatch, outliers, outages or adversarial behavior. To cope with such perturbations, this work introduces a regularized total least-squares (TLS) approach and iterative algorithms with convergence guarantees to solve both tasks. Further generalizations are also considered relying on structured and/or weighted TLS when extra prior information on the perturbation is available. Analyses with simulated and real data corroborate the effectiveness of the novel TLS-based approaches.**

***Index Terms*— Graph and signal perturbations, total least-squares, structural equation models, topology identification, graph signal reconstruction.**

I. Introduction

Graphs play a pivotal role in the analysis of complex systems. In applications such as in biological, financial or social sciences, data-driven graphs are adopted to model (un)directed data dependencies. In physical multiagent systems, graphs are introduced to represent physical or engineered links between vertices of e.g. vehicular, power or communication networks, and they are crucial in tasks such as devising resource allocation strategies or imputing missing data. However, perturbations on links or vertices can be present in both data-induced and physical networks and may compromise the performance of graph-based learning tasks. In a gene regulatory network, for instance, the inferred topology may be imprecise due to e.g., model mismatch or noise in the data; while in a communication network, graph perturbations may arise due to link or node outages.

Recently, the vulnerability of networked systems to failures, anomalies, or model mismatch has received increasing interest [6], [15], [37], [30], [12], [13], [20]. In the context of statistical analysis of network data, error propagation in network characteristics (e.g. count of subgraphs) has been studied in [6] and [15]. In order to account for topological perturbation, probabilistic or uncertain graphs have been considered for clustering [37], graph filtering [30], and consensus [57]. Other works developed tools based on small perturbation analysis of the Laplacian matrix [56] to handle graph perturbations for robust resource allocation [12], graph signal inference [13], and tracking of time-varying graph signals [20]. Signal and graph perturbations via total least squares were first analyzed in our previous work [14], where only preliminary results on synthetic data were studied. Differently from [14], we develop in addition an alternative algorithm for the topology identification and theoretical result for the case of signal recovery.

The present work deals with signal and graph perturbations for the tasks of topology identification and graph signal inference based on total least-squares (TLS). TLS is the generalization of least-squares (LS) tailored to account for error mismatch (a.k.a. noise) present in both the input and the output matrices [54]. TLS and its regularized variants emerge in several applications including system identification [51], information retrieval [31], forecasting of financial data and reconstruction of medical images [41]. Building upon TLS, weighted TLS [4], structured TLS [17], and sparse TLS [58] have also been introduced to incorporate different prior information.

Structural equation models (SEMs) [35] have been widely adopted in diverse fields for network topology identification [5], [11], [24], [25], [43], [45], mostly relying on measurements available across nodes. Topology identification (ID) with partially observed nodal processes has also been studied recently [29], [48]. Leveraging piecewise stationarity, SEMs-based topology inference was pursued in [48] when only (partial) statistics of nodal measurements are given, while a joint inference algorithm was developed in [29] to identify the topology as well as interpolate graph signals based on partial observations of the nodal signals. However, neither of them accounts for signal perturbations. Topology identification approaches that rely on Graphical LASSO and its generalizations have also been developed [23], [34], along with graphical model selection based methods for stationary [33], and non-stationary processes [36], [47]; see also [24], [42]. Different from these approaches, the methods here do not rely on any probabilistic assumptions for the network model, further account for perturbations in the topology or the nodal observations. In contrast, approaches identifying dynamic net-

Work in this paper was supported by grants NSF 1711471, 1500713 and NIH 1R01GM104975-01. Preliminary results of this work was presented in [14]. E. Ceci and S. Barbarossa are with the Department of Information Engineering, Electronics and Telecommunications, Sapienza University of Rome, Rome 00184, Italy (e-mail: elena.ceci@uniroma1.it; sergio.barbarossa@uniroma1.it). Y. Shen is with CPCC and the Department of Electrical Engineering and Computer Science, the University of California, Irvine, 92697, USA (email: yannings@uci.edu). G. B. Giannakis is with the Digital Technology Center and the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: georgios@umn.edu).

work topologies based on vector autoregressive models [10], [50], do not take into account signal perturbations, but only consider additive noise.

Whether adopting parametric [3], [19], [44] or non-parametric approaches [27], [28], [49], most existing works on graph signal reconstruction assume that the graph topologies are known exactly. However, they do not consider that the nominal graph topologies may be inaccurate. Expectation-maximization approaches [18] are used in graphical models to infer iteratively the graph parameters and the missing signals [32], [39], but rely on probabilistic assumptions for nodal signals, which is not the case in the present approach.

The present work relies on TLS and SEMs to cope with two intertwined graph learning tasks, namely:

T1. Topology identification (ID) based on perturbed nodal signal observations; and,

T2. Graph signal inference given partial nodal observations and perturbed topologies.

An example of (T1) would be topology identification of a gene regulatory network from inaccurate data, due to possible errors occurring during the data collection process. As for task (T2), possible deviations from the nominal topology due to break down links need to be taken into account while inferring unknown nodal signals in wireless networks.

The novelties here can be summarized in two directions.

C1. The task of graph topology identification in the presence of signal perturbations is formulated for the first time as a total least-squares problem. Two algorithms with complementary strengths are developed: the first algorithm attains an ε-optimal solution, while the second one can afford a sub-optimal yet computationally efficient solution.

C2. Different from existing methods which deal with signal recovery over a known graph, a novel signal recovery approach is introduced here for graphs with perturbed topologies. Efficient algorithms are then developed to infer the signals, and identifiability of the model is analyzed.

The rest of the paper is organized as follows: Sec. II introduces the context and the TLS formulation with its weighted and structured variants. The topology ID problem (T1) is investigated in Sec. III, while the graph signal inference task (T2) is addressed in Section IV. In Sec. V, synthetic and real data tests are carried out to illustrate the merits of the proposed TLS-based approaches. Finally, concluding remarks and future directions are outlined in Sec. VI.

Notation. Bold lower (upper) case fonts denote column vectors, e.g., $\mathbf{a}$ (matrix $\mathbf{A}$), while operators $(\cdot)^\top$, $\text{vec}(\cdot)$, and $\otimes$ stand for transposition, column-wise matrix vectorization, and Kronecker product, respectively. The $K \times K$ identity matrix is denoted by $\mathbf{I}_K$, and $\mathbf{s}_i$ stands for the i-th canonical vector; while $\text{diag}(\cdot)$, and $\text{bdiag}(\cdot)$ correspondingly represent a diagonal matrix and a block diagonal matrix of its arguments. Finally, the ℓ_1, ℓ_2, and Frobenius norms will be denoted by $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_F$, respectively.

II. Preliminaries

The present section reviews linear SEMs and TLS, along with structured and weighted TLS variants.

A. Structural Equation Models

Consider a directed network of N nodes, whose topology is captured by the adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ with entries $a_{ij} := [\mathbf{A}]_{ij}$, and $a_{ij} \neq 0$ if a directed edge from node j to node i is present. Suppose the network represents a complex system, where y_{it} is the measurement at node i at instant t. The output measurement y_{it} in SEMs depends on its single-hop neighbor measurements, and an exogenous input signal x_{it}, that is

$$y_{it} = \sum_{j\neq i} a_{ij} y_{jt} + b_{ii} x_{it}, \qquad t = 1,\dots,T \tag{1}$$

where $b_{ii} > 0$ weighs the exogenous input. Concatenating nodal measurements in vectors $\mathbf{y}_t{:=}[y_{1t},\dots,y_{Nt}]^\top$, and $\mathbf{x}_t{:=}[x_{1t},\dots,x_{Nt}]^\top$ per slot t, the matrix-vector version of (1) can be compactly written as $\mathbf{y}_t = \mathbf{A}\mathbf{y}_t + \mathbf{B}\mathbf{x}_t,\ t = 1,\dots,T$, where $a_{ii} = 0$ and $\mathbf{B} := \text{diag}(b_{11},\dots,b_{NN})$.

Collecting inputs and outputs[1] across T slots, $N \times T$ matrices $\mathbf{X} := [\mathbf{x}_1,\dots,\mathbf{x}_T]$ and $\mathbf{Y} := [\mathbf{y}_1,\dots,\mathbf{y}_T]$ can be formed, to obtain the linear matrix model

$$\mathbf{Y} = \mathbf{A}\mathbf{Y} + \mathbf{B}\mathbf{X}\,. \tag{2}$$

Existing works treat perturbations as additive observation noise to arrive at the SEM, $\mathbf{Y} = \mathbf{A}\mathbf{Y}+\mathbf{B}\mathbf{X}+\mathbf{V}$, where $\mathbf{V} \in \mathbb{R}^{N\times T}$ is the error matrix. Generally, these works aim to estimate $\mathbf{A}$ (and possibly $\mathbf{B}$), when measurements $\mathbf{Y}$ and $\mathbf{X}$ are given, using least-squares (LS) or regularized LS [7], [11]. On the other hand, when $\mathbf{A}$, $\mathbf{B}\mathbf{X}$ (e.g. obtained by historical data) and a subset of entries of $\mathbf{Y}$ are given, it is also possible to interpolate the unobserved nodal signals [29]. Since existing approaches do not consider possible errors in $\mathbf{A}$ or $\mathbf{Y}$, we are motivated to adopt TLS methods to cope with graph signal and topology perturbations that can be possibly present in SEMs. In particular, if $\mathbf{Y}$ is corrupted by noise, the observed data can be written as $\mathbf{Z} = \mathbf{Y} + \mathbf{E}$, and the model is then given by $\mathbf{Z} - \mathbf{E} = \mathbf{A}(\mathbf{Z} - \mathbf{E}) + \mathbf{B}\mathbf{X}$. Using TLS, we wish to infer $\mathbf{A}$. On the other hand, given a perturbed $\mathbf{A}$ and partial noisy nodal observations (subset of noise-corrupted $\mathbf{Y}$), we aim at recovering the graph signal using a TLS-based approach. Before introducing the formulation of these two tasks, we outline basic TLS notions, and its weighed and structured variants in the following subsection.

B. Weighted and structured TLS

TLS considers the perturbed linear system of equations $\boldsymbol{F} = (\boldsymbol{H}+\boldsymbol{P})\boldsymbol{\Theta} - \boldsymbol{\Sigma}$, where $\boldsymbol{F} \in \mathbb{R}^{M\times T}$ denotes the output matrix with $M < T$, $\boldsymbol{H} \in \mathbb{R}^{M\times N}$ the input (or regression) matrix, $\boldsymbol{\Theta} \in \mathbb{R}^{N\times T}$ an unknown matrix of parameters, while $\boldsymbol{\Sigma} \in \mathbb{R}^{M\times T}$ and $\boldsymbol{P} \in \mathbb{R}^{M\times N}$ capture the error matrices. Different from classical LS where $\mathbf{P} = \mathbf{0}$, TLS treats symmetrically the input and the output in the sense that both $\boldsymbol{H}$ and $\boldsymbol{F}$ may have

[1]Causes-effect per node do not have to happen instantaneously, since causes $\{y_{jt}, x_{it}\}$ can occur at the beginning and effect y_{it} at the end of slot t.

errors due to model mismatch, noise, or outliers. Hence, TLS solves the following problem

$$\min_{\boldsymbol{\Theta},\boldsymbol{P},\boldsymbol{\Sigma}} \quad \|[\boldsymbol{P},\boldsymbol{\Sigma}]\|_F^2 \tag{3a}$$
$$\text{s. to} \quad \boldsymbol{F} = (\boldsymbol{H}+\boldsymbol{P})\boldsymbol{\Theta} - \boldsymbol{\Sigma}. \tag{3b}$$

The structured variants of TLS rely on exploiting the structure of input and output matrices, as well as noise statistics, to achieve improved estimation performance. The structure of a matrix in the TLS context is defined as follows [41], [58].

Definition 1. Given a parameter vector $\boldsymbol{\omega} \in \mathbb{R}^{n_\omega}$, the $M \times (N+T)$ data matrix $[\boldsymbol{H},\boldsymbol{F}](\boldsymbol{\omega})$ has a structure $\mathscr{S}(\boldsymbol{\omega})$ characterized by $\boldsymbol{\omega}$, if and only if there is a mapping such that $\boldsymbol{\omega} \in \mathbb{R}^{n_\omega} \to [\boldsymbol{H},\boldsymbol{F}](\boldsymbol{\omega}) := \mathscr{S}(\boldsymbol{\omega}) \in \mathbb{R}^{M\times(N+T)}$.

Note that Definition 1 reduces to the trivial case when $\boldsymbol{\omega} := \text{vec}([\boldsymbol{H},\boldsymbol{F}])$ with dimension $M(N+T)$, which corresponds to the unstructured case. However, when $\boldsymbol{\omega}$ provides a parsimonious representation of the data matrix with $n_\omega \ll M(N+T)$, we can take advantage of the matrices' structure [41]. By introducing the parameter vector $\boldsymbol{\omega}$ and the noise parameter vector $\boldsymbol{\nu} \in \mathbb{R}^{n_\omega}$, such that $\mathscr{S}(\boldsymbol{\omega}+\boldsymbol{\nu}) := [\boldsymbol{H}+\mathbf{P},\boldsymbol{F}+\boldsymbol{\Sigma}](\boldsymbol{\omega}+\boldsymbol{\nu})$, the Frobenius norm $\|[\boldsymbol{P},\boldsymbol{\Sigma}]\|_F^2$ becomes $\|\boldsymbol{\nu}\|_2^2$. The weighted TLS is obtained if prior knowledge about the $\boldsymbol{\nu}$ is incorporated by weighting the norm $\|\boldsymbol{\nu}\|_2^2$ through the $n_\omega \times n_\omega$ positive definite matrix $\mathbf{W}$. Hence, the structured and weighted TLS (SWTLS) cost is expressed as $\boldsymbol{\nu}^\top\mathbf{W}\boldsymbol{\nu}$. Clearly, when $\mathbf{W}=\mathbf{I}$, the SWTLS boils down to a structured-only form. Here, we will adapt the SWTLS approach to recover the graph signal of interest. Specifically, Definition 1 will be used to capture the nonzero patterns of $\mathbf{A}$ in (1) and (2), when we know a priori that the perturbations occur only on nominal edges. The weight matrix on the other hand, will be employed to incorporate possible prior information about link failure probabilities and the variance of observation error variances (see Sec. IV-A).

III. Topology ID with signal perturbations

Outliers and defects in the measuring process lead to perturbed nodal signals. Such perturbation may affect the topology ID performance. Let us rewrite the observation matrix $\mathbf{Y}$ in (2) as $\mathbf{Z}-\mathbf{E}$, where $\mathbf{E}$ is a perturbation matrix. Given $\mathbf{Z}$ and $\mathbf{BX}$, the aim is to find the adjacency matrix $\mathbf{A}$ from the "measurement-perturbed" SEM

$$\mathbf{Z}-\mathbf{E} = \mathbf{A}(\mathbf{Z}-\mathbf{E}) + \mathbf{BX}. \tag{4}$$

The presence of the perturbation that appears in both sides justifies a formulation inspired by TLS method recalled in (3), with the difference that in our model the perturbation of the input and output matrix is exactly the same, i.e. $\mathbf{E} \in \mathbb{R}^{N\times T}$. In most real-world networks, such as social, transportation, and biological networks, the nodes exhibit a few interconnections and the corresponding adjacency matrix is sparse [5], [26]. Thus, accounting for the latter through a sparsity-promoting regularization term, we formulate a regularized TLS-based approach for "measurement-perturbed" SEM (4) (TLS-SEM) given by

$$\{\hat{\mathbf{A}},\hat{\mathbf{E}}\} = \arg\min_{\mathbf{A},\mathbf{E}} \|\mathbf{E}\|_F^2 + \lambda\|\mathbf{A}\|_1 \tag{5a}$$
$$\text{s.to} \quad \mathbf{Z} = \mathbf{A}(\mathbf{Z}-\mathbf{E}) + \mathbf{BX} + \mathbf{E} \tag{5b}$$
$$a_{ii} = 0, \quad i=1,\ldots,N \tag{5c}$$

where $\lambda > 0$ is the regularization parameter, and constraint (5c) enforces the absence of self-loops in $\mathbf{A}$. Clearly, the optimization problem in (5) is nonconvex. The ensuing subsections will develop two solvers with complementary merits.

A. Bisection-based ε-optimal algorithm

In this subsection, we will first recast (5) into a fractional form that can be solved using a bisection-based (BB) iteration, which is convergent to an ε-optimal solution in a finite number of iterations, even though (5) is nonconvex [8]. The following lemma shows how to reformulate (5) in a fractional form.

Lemma 1. *With $\boldsymbol{\Phi} := \mathbf{Z}-\mathbf{BX}$, and $\boldsymbol{\varphi}_i^\top$ denoting its i-th row, the TLS problem in* (5) *is equivalent to the fractional problem*

$$\hat{\mathbf{A}} = \arg\min_{\{\mathbf{a}_{-i}\}_{i=1}^N} \sum_{i=1}^N \Big[\frac{\|\boldsymbol{\varphi}_i - (\mathbf{Z}_{-i})^\top\mathbf{a}_{-i}\|_2^2}{1+N\|\mathbf{a}_{-i}\|_2^2} + \lambda\|\mathbf{a}_{-i}\|_1\Big] \tag{6}$$

where $\mathbf{a}_{-i}^\top$ is the i-th row of $\mathbf{A}$ without the i-th entry, and $\mathbf{Z}_{-i}$ the $(N-1)\times T$ submatrix of $\mathbf{Z}$ after removing its i-th row.

Proof. Clearly, (5) can be rewritten as

$$\arg\min_{\{\mathbf{a}_i,\boldsymbol{\epsilon}_i\}_{i=1}^N} \quad \sum_{i=1}^N \Big(\frac{1}{N}\Big\|[\mathbf{E}^\top,\sqrt{N}\boldsymbol{\epsilon}_i]\Big\|_F^2 + \lambda\|\mathbf{a}_i\|_1\Big) \tag{7a}$$
$$\text{s. to} \quad \boldsymbol{z}_i = (\mathbf{Z}^\top - \mathbf{E}^\top)\mathbf{a}_i + b_{ii}\mathbf{x}_i + \boldsymbol{\epsilon}_i, \forall i \tag{7b}$$
$$a_{ii} = 0, \forall i \tag{7c}$$

where $\mathbf{a}_i^\top$, $\boldsymbol{z}_i^\top$, $\mathbf{x}_i^\top$, $\boldsymbol{\epsilon}_i^\top$ are the i-th rows of $\mathbf{A}$, $\mathbf{Z}$, $\mathbf{X}$, $\mathbf{E}$, respectively, and b_{ii} is the i-th diagonal entry of $\mathbf{B}$. Thus, the constraint (7b) becomes

$$\boldsymbol{\varphi}_i = (\mathbf{Z}^\top - \mathbf{E}^\top)\mathbf{a}_i + \boldsymbol{\epsilon}_i \,. \tag{8}$$

Next, with $\boldsymbol{v}_i := \text{vec}([\mathbf{E}^\top,\sqrt{N}\boldsymbol{\epsilon}_i])$, we have $\Big\|[\mathbf{E}^\top,\sqrt{N}\boldsymbol{\epsilon}_i]\Big\|_F^2 = \|\boldsymbol{v}_i\|_2^2$; and upon defining $\mathbf{G}(\mathbf{a}_i) := ([-\mathbf{a}_i^\top, \frac{1}{\sqrt{N}}] \otimes \mathbf{I}_T)$, constraint (7b) is re-expressed as

$$\boldsymbol{\varphi}_i - \mathbf{Z}^\top\mathbf{a}_i = \mathbf{G}(\mathbf{a}_i)\boldsymbol{v}_i, \forall i. \tag{9}$$

Note that, with $\mathbf{A}$ fixed, (7) becomes $\min_{\boldsymbol{v}_i} \|\boldsymbol{v}_i\|_2^2$ subject to (9), which admits a closed-form solution

$$\begin{aligned}\boldsymbol{v}_i &= \mathbf{G}^\top(\mathbf{a}_i)[\mathbf{G}(\mathbf{a}_i)\mathbf{G}^\top(\mathbf{a}_i)]^{-1}(\boldsymbol{\varphi}_i - \mathbf{Z}^\top\mathbf{a}_i)\\ &= (\|\mathbf{a}_i\|_2^2 + \frac{1}{N})^{-1}\mathbf{G}^\top(\mathbf{a}_i)(\boldsymbol{\varphi}_i - \mathbf{Z}^\top\mathbf{a}_i)\end{aligned} \tag{10}$$

where the second equality holds because $\mathbf{G}(\mathbf{a}_i)\mathbf{G}^\top(\mathbf{a}_i) = ([-\mathbf{a}_i^\top, \frac{1}{\sqrt{N}}] \otimes \mathbf{I}_T)([-\mathbf{a}_i^\top, \frac{1}{\sqrt{N}}]^\top \otimes \mathbf{I}_T) = (\|\mathbf{a}_i\|_2^2 + \frac{1}{N})\mathbf{I}_T$. Substituting (10) into (7a), and incorporating the constraint (7c), yields (6). $\square$

The fractional problem (6) is separable across rows of $\mathbf{A}$ as

$$\hat{\mathbf{a}}_{-i} = \arg\min_{\mathbf{a}_{-i}} \frac{\left\|\boldsymbol{\varphi}_i - (\mathbf{Z}_{-i})^\top \mathbf{a}_{-i}\right\|_2^2}{1 + N\left\|\mathbf{a}_{-i}\right\|_2^2} + \lambda \left\|\mathbf{a}_{-i}\right\|_1 \tag{11}$$

which can be viewed as a Lagrangian function. Considering the solution $\hat{\mathbf{a}}_{-i}$ for a given multiplier $\lambda > 0$ and letting $\mu := \|\hat{\mathbf{a}}_{-i}\|_1$, (11) is equivalent to

$$\hat{\mathbf{a}}_{-i} = \arg \min_{\mathbf{a}_{-i} \in \boldsymbol{\chi}(\mu)} f(\mathbf{a}_{-i})$$
$$f(\mathbf{a}_{-i}) := \frac{\left\|\boldsymbol{\varphi}_i - (\mathbf{Z}_{-i})^\top \mathbf{a}_{-i}\right\|_2^2}{1 + N\left\|\mathbf{a}_{-i}\right\|_2^2} \tag{12}$$

where $\boldsymbol{\chi}(\mu) := \{\mathbf{a}_{-i} \in \mathbb{R}^{(N-1)} : \|\mathbf{a}_{-i}\|_1 \leq \mu\}$, and the relationship between μ and λ is data dependent.

The fractional problem (6) remains nonconvex, and will be solved using an iterative solver. The solver consists of an outer loop based on bisection [21], and an inner loop using the branch-and-bound method [2]. In the i-th iteration, the outer loop confines the minimum cost in (12) between a lower and an upper bound. These bounds are obtained through the inner iteration, where a surrogate quadratic function is minimized. The surrogate quadratic function has non-fractional form, whose optimization is more convenient than directly optimizing $f(\mathbf{a}_{-i})$. Specifically, with q denoting a given upper bound of the cost in (12), we have

$$0 \leq q^* := \min_{\mathbf{a}_{-i} \in \boldsymbol{\chi}(\mu)} f(\mathbf{a}_{-i}) \leq q. \tag{13}$$

Then, we define

$$g^*(q) := \min_{\mathbf{a}_{-i} \in \boldsymbol{\chi}(\mu)} g(\mathbf{a}_{-i}, q) \tag{14}$$

with $g(\mathbf{a}_{-i}, q) := \|\boldsymbol{\varphi}_i - (\mathbf{Z}_{-i})^\top \mathbf{a}_{-i}\|_2^2 - q\big(1 + N\|\mathbf{a}_{-i}\|_2^2\big)$. Due to (13) and (14), it holds that

$$g^*(q) \leq 0. \tag{15}$$

Let q^* belong to a known interval $\mathcal{I}_i := [l_i, u_i]$ after the i-th outer iteration. Such an interval decreases at every step of the outer loop, and l_i, u_i are chosen depending on the sign of $g(\mathbf{a}_{-i}, q)$ (cf. Alg. 1). In particular, suppose that $g^*(q)$ is obtained at the middle point of $\mathcal{I}_i$, namely $q_m = (u_i + l_i)/2$. The sign of $g(q_m)$ indicates whether (13) holds or not. If $g(q_m) > 0$, then we deduce from (13) that $q^* > q_m > l_i$, and $q^* \in \mathcal{I}_{i+1} := [q_m, u_i]$. On the other hand, $g(q_m) < 0$ implies $q^* \in \mathcal{I}_{i+1} := [l_i, q_m]$. In both cases, the interval at iteration $i + 1$ shrinks through bisection.

Note that, the Hessian of $g(\mathbf{a}_{-i}, q)$ is $\mathbf{H} := 2(\mathbf{Z}_{-i}(\mathbf{Z}_{-i})^\top - qN\mathbf{I})$, and since qN is positive, $\mathbf{H}$ is not guaranteed to be positive or negative definite. Thus, $g(\mathbf{a}_{-i}, q)$ is an indefinite quadratic.

The inner loop employs a branch-and-bound algorithm to find a feasible and δ-optimal solution $\mathbf{a}^*_{\delta,i}$ of (14), such that $g^*(q) \leq g(\mathbf{a}^*_{\delta,i}, q) \leq g^*(q) + \delta$, where δ denotes a specified margin. The branch-and-bound scheme, summarized in Alg. 2, searches for the upper and lower bounds of the function

$$g_{box}(\mathbf{a}_{-i}) = \min_{\mathbf{a}_{-i} \in \boldsymbol{\chi}(\mu), \boldsymbol{a}_L \leq \mathbf{a}_{-i} \leq \boldsymbol{a}_U} g(\mathbf{a}_{-i}, q) \tag{16}$$

Algorithm 1 Bisection-based (BB) scheme

Input : $\boldsymbol{\Phi}$, $\boldsymbol{Z}$, ε, and δ
Output: $\mathbf{A}^*_{\varepsilon,i}$ (ε-optimal solution)
for $i = 1, \ldots, N$ **do**
  Set $l_0 = 0$, $u_0 = \|\boldsymbol{\varphi}_i\|_2^2$, iteration index $j = 0$, achievable cost $f_m = u_0$, and $(\mathbf{a}^*_{\varepsilon,i})^\top = 0$ be the i-th row of $\mathbf{A}^*_{\varepsilon,i}$
  while $u_j - l_j > \varepsilon$ **do**
    Let $q = \frac{u_j + l_j}{2}$
    Call Algorithm 2 and obtain $\mathbf{a}^*_{\delta,i}$
    $f_g = f(\mathbf{a}^*_{\delta,i})$ and $j = j + 1$
    $f_m = f_g$ and $\mathbf{a}^*_{\varepsilon,i} = \mathbf{a}^*_{\delta,i}$ **if** $f_g < f_m$
    Upper and lower bounds updates:
    $u_j = q, l_j = l_{j-1}$ **if** $g(\mathbf{a}^*_{\delta,i}, q) \leq 0$
    $l_j = q, u_j = u_{j-1}$ **if** $g(\mathbf{a}^*_{\delta,i}, q)l \geq \delta$
    $l_j = q - \delta, u_j = u_{j-1}$ **if** $0 < g(\mathbf{a}^*_{\delta,i}, q) < \delta$
    Set $u_j = \min(u_j, f_g)$
  end
end

where the constraint $\boldsymbol{a}_L \leq \mathbf{a}_{-i} \leq \boldsymbol{a}_U$ represents a box that shrinks as iterations progress. The upper bound U of $g_{box}(\mathbf{a}_{-i})$ can be obtained by a sub-optimal yet efficient solver for (16), see e.g., [9], [55]. While the lower bound L of $g_{box}(\mathbf{a}_{-i})$ can be found by minimizing its convex approximation

$$g_L(\mathbf{a}_{-i}, q) = g(\mathbf{a}_{-i}, q) + (\mathbf{a}_{-i} - \boldsymbol{a}_L)^\top \mathbf{D}(\mathbf{a}_{-i} - \boldsymbol{a}_U) \tag{17}$$

where $\mathbf{D}$ is a diagonal positive semi-definite matrix chosen to ensure the convexity of $g_L(\mathbf{a}_{-i}, q)$, as the solution of the following semi-definite program

$$\min_{\mathbf{D}} \quad (\boldsymbol{a}_U - \boldsymbol{a}_L)^\top \mathbf{D}(\boldsymbol{a}_U - \boldsymbol{a}_L) \tag{18a}$$
$$\text{s. to} \quad \mathbf{H} + 2\mathbf{D} \succeq 0 \tag{18b}$$

where (18b) assures (18) to be convex. At each iteration of the inner loop, the initial box constraint of (16) is split depending on how $U - L$ compares with the preselected δ. This splitting process leads to a smaller U and a tighter L. The detailed inner loop is listed in Alg. 2.

In summary, Alg. 2 is called by Alg. 1 to find the δ-optimal solution and evaluate the sign of $g^*(q)$. However, since $\mathbf{a}^*_{\delta,i}$ is δ-optimal, meaning $g^*(q) > g(\mathbf{a}^*_{\delta,i}, q) - \delta$, if $g(\mathbf{a}^*_{\delta,i}, q) > \delta$, we set the lower bound l_{i+1} to q_m; otherwise, if $0 < g(\mathbf{a}^*_{\delta,i}, q) < \delta$ we set $l_{i+1} = q_m - \delta$. As far as convergence is concerned, the following can be established.

Proposition 1. *After at most* $\left\lceil \ln(\frac{\|\boldsymbol{\varphi}_i\|_2^2}{\varepsilon - 2\delta}) / \ln(2) \right\rceil$ *iterations, with* $\varepsilon > 2\delta$*, an* ε*-optimal solution* $\mathbf{a}^*_{\varepsilon,i}$ *to* (13) *is reached, satisfying*

$$\mathbf{a}^*_{\varepsilon,i} \in \boldsymbol{\chi}(\mu), \text{ and } q^* \leq f(\mathbf{a}^*_{\varepsilon,i}) \leq q^* + \varepsilon, \; i = 1, \ldots, N. \tag{19}$$

Proof. See [58]. ☐

This proposition quantifies the number of outer iterations needed by Algorithm 1 to achieve the ε-optimal solution.

Algorithm 2 Branch-and-Bound scheme

Input : $\boldsymbol{\psi}_i$, $\boldsymbol{Z}_i$, q, and δ.
Output: $\mathbf{a}^*_{\delta,i}$ (δ-optimal solution)
Initialize $\boldsymbol{a}_L, \boldsymbol{a}_U, U = \infty, L = -\infty$, and $\mathcal{K} = \{\boldsymbol{a}_L, \boldsymbol{a}_U, L\}$
while $\mathcal{K} \neq \emptyset$ **do**
  Solve (16) to obtain $\hat{\mathbf{a}}^*_{-i}$
  if $g(\hat{\mathbf{a}}^*_{-i}, q) < U$ **then**
    $U = g(\hat{\mathbf{a}}^*_{-i}, q)$ and $\mathbf{a}^*_{\delta,i} = \hat{\mathbf{a}}^*_{-i}$
  end
  Find $\mathbf{D}$ via (18), and find $\check{\mathbf{a}}^*_{-i}$ and $L = g_L(\check{\mathbf{a}}^*_{-i})$ minimizing (17).
  if $U - L > \delta$ {*split*} **then**
    Find $k = \max_n([\boldsymbol{a}_U]_n - [\boldsymbol{a}_L]_n)$
    Set $\boldsymbol{a}_{L,1} = \boldsymbol{a}_L(\boldsymbol{a}_{U,1} = \boldsymbol{a}_U)$ and $\boldsymbol{a}_{L,2} = \boldsymbol{a}_L(\boldsymbol{a}_{U,2} = \boldsymbol{a}_U)$ except the k-th entry:
    $[\boldsymbol{a}_{L,1}]_k = [\boldsymbol{a}_L]_k$ and $[\boldsymbol{a}_{U,1}]_k = \frac{[\boldsymbol{a}_U]_k + [\boldsymbol{a}_L]_k}{2}$
    $[\boldsymbol{a}_{U,2}]_k = [\boldsymbol{a}_U]_k$ and $[\boldsymbol{a}_{L,2}]_k = \frac{[\boldsymbol{a}_U]_k + [\boldsymbol{a}_L]_k}{2}$.
    Compute $\mathbf{D}_1$, $\mathbf{D}_2$ and L_1, L_2 for each new boxes and $\mathcal{K} = (\boldsymbol{a}_{L,1}, \boldsymbol{a}_{U,1}, L_1) \cup (\boldsymbol{a}_{L,2}, \boldsymbol{a}_{U,2}, L_2)$
    Compare L_1 and L_2 with U:
    $\mathcal{K} = \mathcal{K} \setminus (\boldsymbol{a}_{L,1}, \boldsymbol{a}_{U,1}, L_1)$, **if** $L_1 > U$
    $\mathcal{K} = \mathcal{K} \setminus (\boldsymbol{a}_{L,2}, \boldsymbol{a}_{U,2}, L_2)$, **if** $L_2 > U$
    $\mathcal{K} = \mathcal{K} \setminus (\boldsymbol{a}_{L,m}, \boldsymbol{a}_{U,m}, L_m)$, $m = \arg\min(L_m)$, **o.w.**
  else
    $\mathcal{K} = \mathcal{K} \setminus (\boldsymbol{a}_L, \boldsymbol{a}_U, L)$
  end
end

B. Alternating descent algorithm

The bisection-based solver developed in the previous subsection can approach the global optimum of the fractional TLS, but it is computationally demanding. This prompts the efficient alternative we introduce next with guaranteed convergence at least to a stationary point. We reformulate (5), substituting (5b) into (5a), and we add $\|\mathbf{E}\|_F^2$ to the cost function to constraint the error norm to be small, obtaining

$$\{\hat{\mathbf{A}}, \hat{\mathbf{E}}\} = \arg\min_{\mathbf{A},\mathbf{E}} \|\mathbf{E}\|_F^2 + \|\mathbf{Z} - \mathbf{A}(\mathbf{Z} - \mathbf{E}) - \mathbf{B}\mathbf{X}\|_F^2 + \lambda \|\mathbf{A}\|_1 \tag{20a}$$

$$\text{s. to} \quad a_{ii} = 0, \; i = 1, \ldots, N. \tag{20b}$$

Note that the minimization of (20) does not guarantee that (5b) is still satisfied. Problem (20) is convex with respect to (wrt) each block (matrix) variable $\mathbf{A}$ and $\mathbf{E}$. This motivates an alternating descent iteration to find a sub-optimal yet efficient solution. At iteration $k+1$, given $\hat{\mathbf{A}}[k]$, the error matrix can be estimated as

$$\hat{\mathbf{E}}[k+1] = \arg\min_{\mathbf{E}} \|\mathbf{Z} - \hat{\mathbf{A}}[k](\mathbf{Z} - \mathbf{E}) - \mathbf{B}\mathbf{X}\|_F^2 + \|\mathbf{E}\|_F^2 \tag{21}$$

which admits the closed-form solution

$$\hat{\mathbf{E}}[k+1] = (\hat{\mathbf{A}}^\top[k]\hat{\mathbf{A}}[k] + \mathbf{I}_N)^{-1}\hat{\mathbf{A}}^\top[k](\hat{\mathbf{A}}[k]\mathbf{Z} + \mathbf{B}\mathbf{X} - \mathbf{Z}). \tag{22}$$

Likewise, given $\hat{\mathbf{E}}[k+1]$, the adjacency matrix is updated as

$$\hat{\mathbf{A}}[k+1] = \arg\min_{\mathbf{A}} \|\mathbf{Z} - \mathbf{A}(\mathbf{Z} - \hat{\mathbf{E}}[k+1]) - \mathbf{B}\mathbf{X}\|_F^2 + \lambda \|\mathbf{A}\|_1 \tag{23}$$

which is strongly convex and can be solved via proximal gradient iterations reaching the global optimum. The derivation of the algorithm is omitted here, see [5] for details.

As far as computations, the operation in (22) incurs complexity $O(N^2T)$, when $N \leq T$, while in the worst case the minimum of (23) can be reached in $O(1/\varepsilon)$ iterations; or, $O(1/\sqrt{\varepsilon})$ using fast iterative shrinkage-thresholding algorithms, where ε is the precision of the solution, and each row of $\mathbf{A}$ can be updated in parallel; see [5]. Specifically, the proximal gradient algorithms entail matrix-vector multiplication and soft thresholding operations per row of $\mathbf{A}$. If the number of iterations needed for the proximal gradient algorithm to converge is relatively smaller than N (as we observed in our numerical tests), these operations are negligible when compared to $O(N^2T)$ of (22).

In addition, if $\mathbf{B}$ is also unknown, it can be treated as a variable and estimated along with the rest. In this case, problem (20) is still per-block convex, and $\mathbf{B}$ can be readily found as in [5]. Under regularity conditions the alternating minimization method is guaranteed to converge at least to a stationary point, as asserted in the following proposition.

Proposition 2. *The iterates in* (22) *and* (23) *converge monotonically at least to a stationary point of problem* (20)*.*

Proof. See [53]. □

C. Topology ID with sparse signal perturbations

So far we have seen perturbations affecting all nodal measurements. In a number of settings, however, only a small subset of nodes can be influenced. For instance, in a heterogeneous network, some devices, e.g. sensors, may be less reliable than others. In this case, sparsity of the signal perturbations is well motivated. Introducing a sparse regularizer yields the sparse TLS (sparseTLS) SEM

$$\{\hat{\mathbf{A}}, \hat{\mathbf{E}}\} = \arg\min_{\mathbf{A},\mathbf{E}} \|\mathbf{Z} - \mathbf{A}(\mathbf{Z} - \mathbf{E}) - \mathbf{B}\mathbf{X}\|_F^2 + \lambda_E \|\mathbf{E}\|_1 + \lambda_A \|\mathbf{A}\|_1 \tag{24a}$$

$$\text{s. to} \quad a_{ii} = 0, \; i = 1, \ldots, N \tag{24b}$$

where $\lambda_A > 0$ and $\lambda_E > 0$ are sparsity promoting scalars.

In certain applications such as sensor networks, we may even know which nodes are the more sensitive or vulnerable, which prompts us to leverage additional structure, namely the nonzero pattern of the error matrix. Hence, we write $\mathbf{E}$ as

$$\mathbf{E} = \sum_{e=1}^{N_E} \upsilon_e(\mathbf{n}_e \cdot \mathbf{t}_e^\top) \tag{25}$$

where $\boldsymbol{\upsilon} := [\upsilon_1, \ldots, \upsilon_{N_E}]^\top$ is the collection of the nonzero values of $\text{vec}(\mathbf{E}^\top)$; the $N \times 1$ vector $\mathbf{n}_e$ has all zero entries except one that equals unity in the node affected by the e-th error value; and, $\mathbf{t}_e$ is the $T \times 1$ vector of all zeros except one that equals unity in the observation instant of the e-th error value. The structured error (s)TLS-SEM is then formulated as

$$\begin{aligned}\{\hat{\mathbf{A}}, \hat{\boldsymbol{\upsilon}}\} = \arg\min_{\mathbf{A},\boldsymbol{\upsilon}} & \|\mathbf{Z} - \mathbf{A}(\mathbf{Z} - \sum_{e=1}^{N_E} \upsilon_e(\mathbf{n}_e \cdot \mathbf{t}_e^\top)) - \mathbf{B}\mathbf{X}\|_F^2 \\ & + \lambda_E \|\boldsymbol{\upsilon}\|_2^2 + \lambda_A \|\mathbf{A}\|_1 \\ \text{s. to} \quad & a_{ii} = 0, \forall i \end{aligned} \tag{26}$$

where $\lambda_E > 0$ and $\lambda_A > 0$. The sTLS-SEM problem is still per-block convex, but can be solved by alternating minimization, as in the previous subsection.

IV. Signal inference with topology perturbations

Besides topology ID, another problem that oftentimes arises in graph-related applications is graph signal inference. In many cases, signals over all the nodes may not be available, due to, e.g., energy-saving or privacy reasons. Hence, it is necessary to reconstruct the signal over the unobserved nodes, given the graph topologies. However, such topologies may be perturbed, due to, e.g., link outages, in communication or power networks. This motivates the goal of this section to recover $\mathbf{Y}$, given a possibly perturbed adjacency matrix and the signal observed over a subset of nodes, indexed by $\mathcal{S}_t$ at each instant t. The observation model can then be written as

$$\boldsymbol{\psi}_t = \mathbf{D}_{\mathcal{S}_t}(\mathbf{y}_t + \boldsymbol{\varepsilon}_t)\,, \quad t = 1, \ldots, T \tag{27}$$

where $\mathbf{D}_{\mathcal{S}_t} := \text{diag}(d_{11}^{(t)}, \ldots, d_{NN}^{(t)})$, and $d_{ii}^{(t)} = 1$ if $i \in \mathcal{S}_t$, and zero otherwise; $\boldsymbol{\varepsilon}_t \in \mathbb{R}^N$ denotes the observation error; and, $\boldsymbol{\psi}_t \in \mathbb{R}^N$ represents the observation at time t, with $|\mathcal{S}_t| := M < N$ nonzero entries. For simplicity in exposition, M is considered fixed over time, but it can be generalized as time-varying.

With $\mathbf{A}_0$ denoting the given nominal adjacency matrix, and $\boldsymbol{\Delta} \in \mathbb{R}^{N\times N}$ the topology perturbation matrix, the linear SEM in (2) becomes

$$\mathbf{Y} = (\mathbf{A}_0 - \boldsymbol{\Delta})\mathbf{Y} + \mathbf{B}\mathbf{X} \tag{28}$$

where $\mathbf{A}_0 - \boldsymbol{\Delta}$ is the perturbed adjacency matrix. As in the previous section, we consider $\mathbf{B}\mathbf{X}$ given, e.g. acquired from historical data or $\mathbf{B}\mathbf{X} = 0$ when $\mathbf{X}$ is not present, since the focus of the present section is to identify $\boldsymbol{\Delta}$ and $\{\mathbf{y}_t\}_{t=1}^T$. Resorting to TLS to account for topology perturbations, the topology perturbation aware TLS-SEM can be written as (cf. (27) and (28))

$$\begin{aligned}\{\hat{\boldsymbol{\Delta}}, \hat{\mathbf{Y}}\} = \arg\min_{\boldsymbol{\Delta},\mathbf{Y}} & \lambda_1\|\boldsymbol{\Delta}\|_1 + \lambda_2 \sum_{t=1}^T \|\boldsymbol{\psi}_t - \mathbf{D}_{\mathcal{S}_t}\mathbf{y}_t\|_2^2 \\ & + \|\mathbf{Y} - (\mathbf{A}_0 - \boldsymbol{\Delta})\mathbf{Y} - \mathbf{B}\mathbf{X}\|_F^2 \qquad \text{(29a)} \\ \text{s.to} \quad & [\boldsymbol{\Delta}]_{ii} = 0, \quad i = 1, \ldots, N \qquad \text{(29b)}\end{aligned}$$

where the ℓ_1-norm promotes sparsity of the perturbed links. In addition to sparsity, it has been shown that the elastic net regularizer [59] leads to improved recovery when the network weights are highly correlated [52]. Motivated by this, the elastic norm regularized TLS (elTLS) approach to signal recovery yields

$$\begin{aligned}\{\hat{\boldsymbol{\Delta}}, \hat{\mathbf{Y}}\} = \arg\min_{\boldsymbol{\Delta},\mathbf{Y}} & \sum_{t=1}^T \|\boldsymbol{\psi}_t - \mathbf{D}_{\mathcal{S}_t}\mathbf{y}_t\|_2^2 + \lambda_{1\Delta}\|\boldsymbol{\Delta}\|_1 \\ & + \lambda_{2\Delta}\|\boldsymbol{\Delta}\|_F^2 + \lambda_Y \|\mathbf{Y} - (\mathbf{A}_0 - \boldsymbol{\Delta})\mathbf{Y} - \mathbf{B}\mathbf{X}\|_F^2 \\ \text{s. to} \quad & [\boldsymbol{\Delta}]_{ii} = 0, \quad i = 1, \ldots, N \end{aligned} \tag{30}$$

where $\lambda_{1\Delta} > 0$, $\lambda_{2\Delta} > 0$, and $\lambda_Y > 0$.

The costs in (29) and (30) are both per-block convex, and can be solved iteratively via alternating minimization with guaranteed convergence to at least a stationary point, as argued in Proposition 2.

A. Structured and weighted TLS under topology perturbations

In this subsection, we exploit the structure of the nominal adjacency matrix along with prior information on the perturbations. The goal here is to formulate a structured and weighted TLS problem (cf. Sec. II-B) for the signal inference task under topology perturbations. Denoting with L the number of links of the nominal graph and $\boldsymbol{\omega} := [\omega_1, \ldots, \omega_L]^\top$ the vector collecting the nonzero edge weights, the nominal adjacency matrix can be represented as (cf. Definition 1)

$$\mathbf{A}_0 = \mathscr{S}(\boldsymbol{\omega}) := \sum_{l=1}^L \omega_l(\mathbf{s}_{u_l}\mathbf{s}_{v_l}^\top) \tag{31}$$

where (u_l, v_l) are the incident nodes of link l, and $\mathbf{s}_i$ the $N\times 1$ i-th canonical vector. The structure $\mathscr{S}(\boldsymbol{\omega})$ accounts for the L nonzero entries of $\mathbf{A}_0$. Assuming that perturbations occur only on the existing links, it will also allow us to reduce the number of unknown perturbations from N^2 to L.

According to Sec. II-B and (31), we will parameterize $\mathbf{A}_0$ using $\boldsymbol{\omega}$, and correspondingly $\boldsymbol{\Delta}$ via $\boldsymbol{\nu} := [\nu_1, \ldots, \nu_L]^\top$, whose nonzero entries represent a failure or error in the edge weight. Thus, the perturbed adjacency matrix is given by

$$\mathbf{A}_0 - \boldsymbol{\Delta} = \mathscr{S}(\boldsymbol{\omega} - \boldsymbol{\nu}) := \sum_{l=1}^L (\omega_l - \nu_l)(\mathbf{s}_{u_l}\mathbf{s}_{v_l}^\top). \tag{32}$$

In some cases, extra information such as the link failure probabilities $\{\pi_l\}_{l=1}^L$ and the observation noise variance $\{\sigma_i^2\}_{i=1}^N$ can be available across nodes. Such prior information can be collected after observing the network over time and recording the occurrence of failures, as well as the statistics of the measurement errors.

Let $\mathbf{W}_A := \text{diag}(r(\pi_1) \ldots r(\pi_L))$ denote the topology reliability weight matrix, where $r(\pi_l)$ is a known function of π_l, e.g. $r(\pi_l) = \pi_l^{-1}$, and likewise $\mathbf{W}_\Psi := [\text{diag}(\sigma_1^2 \ldots \sigma_N^2)]^{-1}$ for the measurement errors. In order to use an SWTLS cost (cf. Sec. II-B), we replace the first two terms in (29a) with the weighted ℓ_1-norm of the topology error vector $\|\mathbf{W}_A\boldsymbol{\nu}\|_1$, and the sum of the weighted ℓ_2-norm of the observation errors $\sum_{t=1}^{T} \|\boldsymbol{\psi}_t - \mathbf{D}_{\mathcal{S}_t}\mathbf{y}_t\|^2_{\mathbf{W}_\Psi}$. Combining with (32), the regularized SWTLS-based SEMs can be written as

$$\{\hat{\boldsymbol{\nu}}, \hat{\mathbf{Y}}\} = \arg\min_{\boldsymbol{\nu},\mathbf{Y}} \quad \lambda_1 \|\mathbf{W}_A\boldsymbol{\nu}\|_1 + \lambda_2 \sum_{t=1}^{T} \|\boldsymbol{\psi}_t - \mathbf{D}_{\mathcal{S}_t}\mathbf{y}_t\|^2_{\mathbf{W}_\Psi} + \|\mathbf{Y} - \sum_{l=1}^{L}(\omega_l - \nu_l)(\mathbf{s}_{u_l}\mathbf{s}_{v_l}^\top)\mathbf{Y} - \mathbf{B}\mathbf{X}\|_F^2 \tag{33}$$

which can be solved via alternating minimization. Given $\hat{\boldsymbol{\nu}}[k]$ from iteration k, and exploiting the separability across columns of $\mathbf{Y}$, the graph signal at $k+1$ is reconstructed per slot t as

$$\hat{\mathbf{y}}_t[k+1] = \arg\min_{\mathbf{y}_t} \lambda_2 \|\boldsymbol{\psi}_t - \mathbf{D}_{\mathcal{S}_t}\mathbf{y}_t\|^2_{\mathbf{W}_\Psi} + \|\mathbf{y}_t - \sum_{l=1}^{L}(\omega_l - \hat{\nu}_l[k])y_{v_l,t}\mathbf{s}_{u_l} - \mathbf{B}\mathbf{x}_t\|_2^2 \tag{34}$$

where $\mathbf{s}_{v_l}^\top\mathbf{y}_t = y_{v_l,t}$ because $\mathbf{s}_{v_l}$ is a canonical vector.

The minimization in (34) leads to the closed-form update

$$\hat{\mathbf{y}}_t[k+1] = (\mathbf{C}^\top[k]\mathbf{C}[k] + \lambda_2\mathbf{D}_{\mathcal{S}_t}^\top\mathbf{W}_\Psi\mathbf{D}_{\mathcal{S}_t})^{-1}(\mathbf{C}^\top[k]\mathbf{B}\mathbf{x}_t + \lambda_2\mathbf{D}_{\mathcal{S}_t}^\top\mathbf{W}_\Psi\boldsymbol{\psi}_t)\,, \qquad t = 1,\ldots,T \tag{35}$$

with $\mathbf{C}[k] := (\mathbf{I}_N - \sum_{l=1}^{L}(\omega_l - \hat{\nu}_l[k])\mathbf{s}_{u_l}\mathbf{s}_{v_l}^\top)$.

Given $\hat{\mathbf{Y}}[k+1] = [\hat{\mathbf{y}}_1[k+1],\ldots,\hat{\mathbf{y}}_T[k+1]]$, we can exploit in (33) the separability across rows of $\mathbf{Y}$. Let L_n denote the number of neighbors of node n, and $\boldsymbol{\omega}_n := [\omega_1^{(n)},\ldots,\omega_{L_n}^{(n)}]^\top$ and $\boldsymbol{\nu}_n := [\nu_1^{(n)},\ldots,\nu_{L_n}^{(n)}]^\top$ the vectors collecting edge and error weights in the neighborhood of n. Similarly, let the diagonal matrix $\mathbf{W}_A^n$ be the n-th block of the block diagonal matrix $\mathbf{W}_A$. With $\boldsymbol{\gamma}_n^\top$ and $\boldsymbol{x}_n^\top$ representing the n-th row of $\mathbf{Y}$ and $\mathbf{X}$, respectively, $\hat{\boldsymbol{\nu}}_n[k+1]$ can be updated as

$$\hat{\boldsymbol{\nu}}_n[k+1] = \arg\min_{\boldsymbol{\nu}_n} \lambda_1 \|\mathbf{W}_A^n\boldsymbol{\nu}_n\|_1 + \|\hat{\boldsymbol{\gamma}}_n[k+1] - (\hat{\mathbf{Y}}_n[k+1])^\top(\boldsymbol{\omega}_n - \boldsymbol{\nu}_n) - b_{nn}\boldsymbol{x}_n\|_2^2 \tag{36}$$

where $\mathbf{Y}_n$ is a submatrix of $\mathbf{Y}$ formed by the rows corresponding to the neighboring nodes of n in the nominal topology. Sub-problem (36) is again convex, but not differentiable, which suggests an iterative proximal gradient solver.

The complexity of (35) is $O(N^3)$, and estimation can be parallelized across $\mathbf{y}_t$ for $t = 1,\ldots,T$. In the worst case, the minimum of (36) can be reached in $O(1/\varepsilon)$ iterations, or $O(1/\sqrt{\varepsilon})$ using fast iterative shrinkage-thresholding algorithms [5], where ε is the precision of the solution. In addition, all $\{\boldsymbol{\nu}_n\}$ can be computed in parallel. Such proximal gradient solvers entail matrix-vector multiplication and soft thresholding operations, the complexity of which can be negligible relative to $O(N^3)$, when $\{L_n\}$ are much smaller than N.

B. Identifiability of topology perturbations

In this subsection, we investigate conditions that ensure uniqueness in identifying the perturbation vector $\boldsymbol{\nu}$ in the noise-free[2] structured topology perturbation model in Sec. IV-A (cf. (28) and (32)). To this end, consider the n-th row of the $N \times T$ matrix $\mathbf{Y}$ in (28), which can be expressed as

$$\mathbf{y}_n^\top = (\mathbf{a}_n^\top - \boldsymbol{\delta}_n^\top)\mathbf{Y} + b_{nn}\mathbf{x}_n^\top \tag{37}$$

with $\mathbf{a}_n^\top$ and $\boldsymbol{\delta}_n^\top$ likewise denoting the nth rows of $\mathbf{A}_0$ and $\boldsymbol{\Delta}$, respectively. With L_n being the number of neighbors of node n, we define the $1 \times L_n$ vector $\boldsymbol{\omega}_n^\top$ formed after removing the zero entries of $\mathbf{a}_n^\top$ per node n; and similarly the $1 \times L_n$ vector $\boldsymbol{\nu}_n^\top$ after removing the corresponding entries of $\boldsymbol{\delta}_n^\top$. Using these definitions, (37) can be simplified to

$$\mathbf{y}_n^\top = (\boldsymbol{\omega}_n^\top - \boldsymbol{\nu}_n^\top)\mathbf{Y}_n + b_{nn}\mathbf{x}_n^\top \tag{38}$$

where $\mathbf{Y}_n$ is an $L_n \times T$ submatrix obtained after removing the rows of $\mathbf{Y}$ corresponding to the zero entries of $\mathbf{a}_n^\top$.

To take into account the number of samples T_n per node n, we further introduce the $T_n \times T$ matrix $\boldsymbol{\mathcal{D}}_n$ obtained after removing the all-zero rows of the $T \times T$ diagonal matrix $\mathrm{diag}\{d_{nn}^{(1)} \ldots d_{nn}^{(T)}\}$, where $d_{nn}^{(t)} = 1$ if node is sampled at slot t, and $d_{nn}^{(t)} = 0$ otherwise. Multiplying $\boldsymbol{\mathcal{D}}_n$ from the right with a matrix, selects T_n (out of T) rows corresponding to the time-slot indices that node n is sampled. We rely on $\boldsymbol{\mathcal{D}}_n$ to form the $T_n \times 1$ vector $\boldsymbol{\phi}_n := \boldsymbol{\mathcal{D}}_n\mathbf{y}_n$, which after employing the transposed version of (38) can be expressed as

$$\boldsymbol{\phi}_n = \boldsymbol{\mathcal{D}}_n[\mathbf{Y}_n^\top(\boldsymbol{\omega}_n - \boldsymbol{\nu}_n) + b_{nn}\mathbf{x}_n]\,. \tag{39}$$

Motivated by the fact that e.g., adversaries can compromise only a few links per node n, it is reasonable to explore identifiability conditions when the sought perturbation vector $\boldsymbol{\nu}_n$ is sparse with p_n $(< L_n)$ nonzero entries.

Arguing by contradiction to establish that $\boldsymbol{\nu}_n$ can be uniquely identified from (39), we will suppose that there exists another $L_n \times 1$ vector $\boldsymbol{\xi}_n \neq \boldsymbol{\nu}_n$ with p_n nonzero entries satisfying $\boldsymbol{\phi}_n = \boldsymbol{\mathcal{D}}_n[\mathbf{Y}_n^\top(\boldsymbol{\omega}_n - \boldsymbol{\xi}_n) + b_{nn}\mathbf{x}_n]$. Subtracting the latter from (39), yields

$$\mathbf{0} = \boldsymbol{\mathcal{D}}_n\mathbf{Y}_n^\top(\boldsymbol{\nu}_n - \boldsymbol{\xi}_n)\,. \tag{40}$$

Clearly, the difference $\boldsymbol{\nu}_n - \boldsymbol{\xi}_n$ of the two p_n-sparse vectors $\boldsymbol{\nu}_n$ and $\boldsymbol{\xi}_n$, has at most $2p_n$ nonzero entries; and with $p_{\max} := \max_{n=1,\ldots,N} p_n$, we have that the differences $\{\boldsymbol{\nu}_n - \boldsymbol{\xi}_n\}$ across all nodes can have at most $2p_{\max}$ nonzero entries.

To proceed with specifying identifiability conditions of our sparse vector differences, we will need the following definition of the Kruskal rank of a matrix.

Definition 2 [38]. *The Kruskal rank of a matrix* $\mathbf{M}$*, denoted as* $\mathrm{kr}(\mathbf{M})$*, is defined as the maximum number* ρ *such that any combination of* ρ *columns of* $\mathbf{M}$ *constitutes a full-rank submatrix.*

Since the $2p_{\max}$ nonzero entries of $\boldsymbol{\nu}_n - \boldsymbol{\xi}_n$ can occur in *any* subset of this vector difference, we deduce that having $\mathrm{kr}(\boldsymbol{\mathcal{D}}_n\mathbf{Y}_n^\top) \geq 2p_{\max}$, guarantees that any $2p_{\max}$ columns of $\boldsymbol{\mathcal{D}}_n\mathbf{Y}_n^\top$ submatrix will be full rank. Under this condition, we find from (40) that $\boldsymbol{\nu}_n = \boldsymbol{\xi}_n$, which leads to contradiction. Summarizing, we have established the following result.

Proposition 3. *If* $\mathrm{kr}(\boldsymbol{\mathcal{D}}_n\mathbf{Y}_n^\top) \geq 2\max_n p_n$*, the* p_n*-sparse perturbation vector* $\boldsymbol{\nu}_n$ *is identifiable from* (39)*, for* $n = 1,\ldots,N$*.*

Intuitively, Proposition 3 asserts that sparsity in the perturbation renders the bound on the Kruskal rank easier to satisfy,

[2]Absence of noise ($\boldsymbol{\epsilon}_t \equiv \mathbf{0}\ \forall t$) is typically assumed in identifiability studies, in order to isolate (non) uniqueness issues from estimation errors.

and thus ensure identifiability. As a word of caution, it is worth mentioning that finding the Kruskal rank of a matrix is combinatorially complex in its dimensions [38]. In addition, this condition may be impossible to check because matrix $\boldsymbol{\mathcal{D}}_n \mathbf{Y}_n^\top$ is not always observed in practice.

V. Numerical tests

In this section, we present several synthetic and real data tests for the novel TLS-based algorithms, both for topology ID under signal perturbations, and graph signal inference under topology perturbations. The regularization parameters are selected by grid search cross-validation for all the algorithms.

A. Synthetic tests for topology ID under signal perturbations

1) Bisection-based versus alternating descent iterations: For this test, the adjacency matrix $\mathbf{A}^{(0)}$ is simulated as a 6×6 matrix of binary 0-1 entries with 2 nonzero entries per row, and $\mathbf{Z} = \mathbf{Y} + \mathbf{E}$, with $[\mathbf{E}]_{ij} \sim \mathcal{N}(0, 10^{-2})$, while the observation $\mathbf{Y} = (\mathbf{I}_N - \mathbf{A}^{(0)})^{-1}\mathbf{B}\mathbf{X}$, with $\mathbf{B} = \mathbf{I}_N$ and $[\mathbf{X}]_{ij} \sim \mathcal{U}[0, 1.5]$. Alg. 1 is tested with $\mu = 5$, $\boldsymbol{a}_L = \mathbf{0}$, and $\boldsymbol{a}_U = \mathbf{1}$.

Fig. 1 shows the performance reached by the alternating descent (AD) iterations in (22) and (23), the conventional least-squares (LS) SEM [5], [11], and the BB iterations (Subsection III-A), all in terms of $\text{MSE}_A = \sum_{ij} (\hat{a}_{ij} - a_{ij})^2 / N^2$, for different values of ε. The ε-optimal BB solver improves as ε decreases, while the solutions of the AD and LS-SEM schemes do not depend on ε, and hence are constant $\forall \varepsilon$. For $\varepsilon < 10^{-2}$, both perturbation-aware methods outperform the LS-SEM method. Note that the BB method slightly outperforms the AD one. However, the BB algorithm is computationally demanding.

Fig. 2 depicts the runtime of the three competing algorithms in seconds,[3] when $\varepsilon = 10^{-3}$, and for $N = 4, 6,$ and 9. The figure demonstrates that the AD method is computationally more efficient than the BB scheme. For this reason, the following tests will include only the AD iteration, which will be henceforth abbreviated as TLS-SEM.

2) Topology ID under signal perturbations: Here, we test the performance of the AD solver (20) for simulated data, and compare it with LS-SEM. We generated a Kronecker graph with $N = 64$ as in [40], and $\mathbf{B} = \mathbf{I}_N$ was assumed given. We generated random matrices with uniformly distributed entries $[\mathbf{X}]_{it} \sim \mathcal{U}[0, 1.5]$, and Gaussian distributed entries $[\mathbf{E}]_{it} \sim \mathcal{N}(0, \sigma_E^2)$. Matrices $\mathbf{Y}$ and $\mathbf{Z}$ were then constructed according to (2) and (4), with $T = 120$, while λ was selected via cross-validation. Fig. 3 shows the MSE_A performance of LS-SEM and TLS-SEM for different SNR(dB):$= 10\log_{10}(\|\bar{\mathbf{y}}\|_2^2/(N\sigma_E^2))$ and $\bar{\mathbf{y}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{y}_t$. It can be observed that TLS-SEM outperforms LS-SEM. Fig. 4 shows the performance versus different number of samples T, with fixed $\sigma_E = 0.2$. Evidently, TLS-SEM outperforms LS-SEM even when the number of observations is small.

[3]This experiment was run on a machine with i5-6200U @2.30 GHz CPU, and 8GB of RAM.

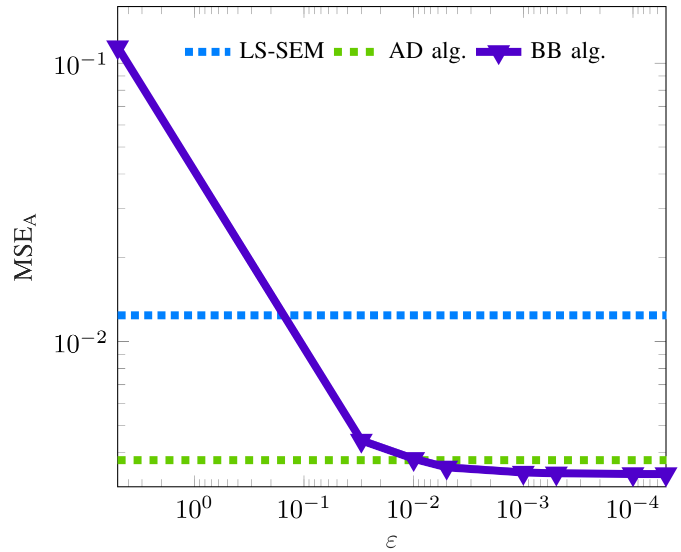

Figure 1: MSE_A across ε, obtained by the ε-optimal algorithm. This result is compared with the LS-SEM, and with the AD (TLS-SEM) iteration.

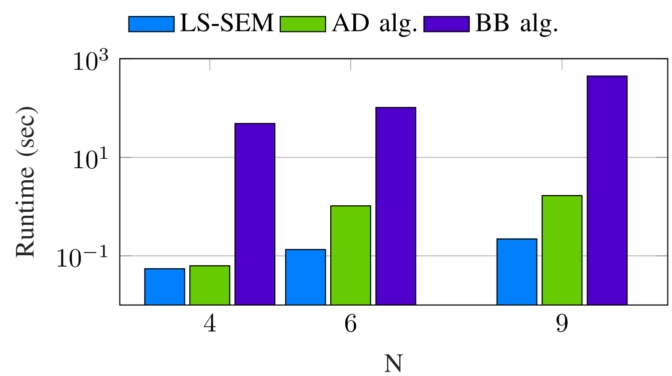

Figure 2: Runtime in seconds.

3) Sparse signal perturbation: In this experiment, we tested the performance of sparse TLS in (24) and (26). We generated an adjacency matrix as a Kronecker graph of size 64×64 with binary entries. Entries of $\mathbf{X}$ were generated as uniform i.i.d. random variables, that is $[\mathbf{X}]_{ij} \sim \mathcal{U}[0, 1.5]$, and $\mathbf{B} = \mathbf{I}_N$. Furthermore, we set $\mathbf{Z} = \mathbf{Y} - \mathbf{E}$, where $\mathbf{Y} = (\mathbf{I}_N - \mathbf{A})^{-1}\mathbf{B}\mathbf{X}$, and the sparse $\mathbf{E}$ was generated such that $\mathbf{E}$ has zero entries on $N_0 = N - 8$ selected rows, while the nonzero entries were drawn from a uniform distribution over $[0, 0.3]$.

Fig. 5 shows the performance of LS-SEM, TLS-SEM in (20), sparseTLS in (24) and sTLS-SEM in (26), in terms of MSE_A for different T. The TLS-SEM methods outperform LS-SEM, and the performance gain increases as more data samples are collected. Results of this subsection were averaged over 100 realizations of $\mathbf{X}$ and $\mathbf{E}$.

B. Real data tests for topology ID with signal perturbations

In this subsection, we present experiments on gene expression data to identify the underlying gene regulatory network. The data were collected from RNA sequencing of cell samples derived from 69 unrelated Nigerian individuals, exten-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2020.2982833, IEEE Transactions on Signal Processing

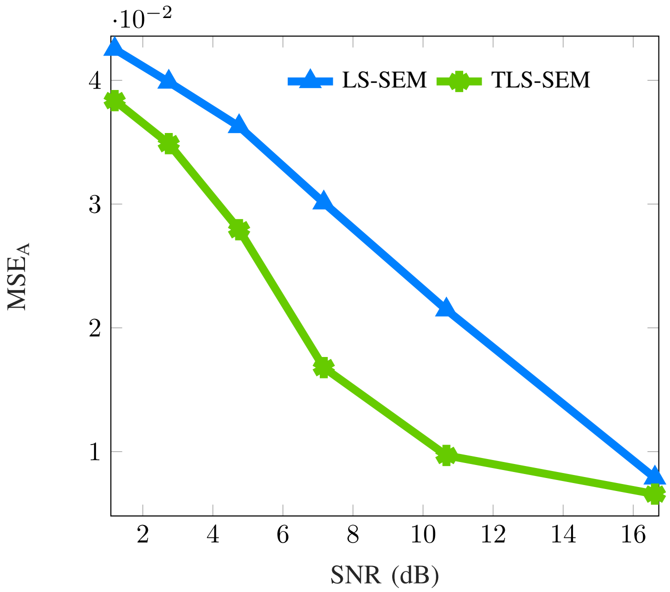

Figure 3: MSE_A versus SNR.

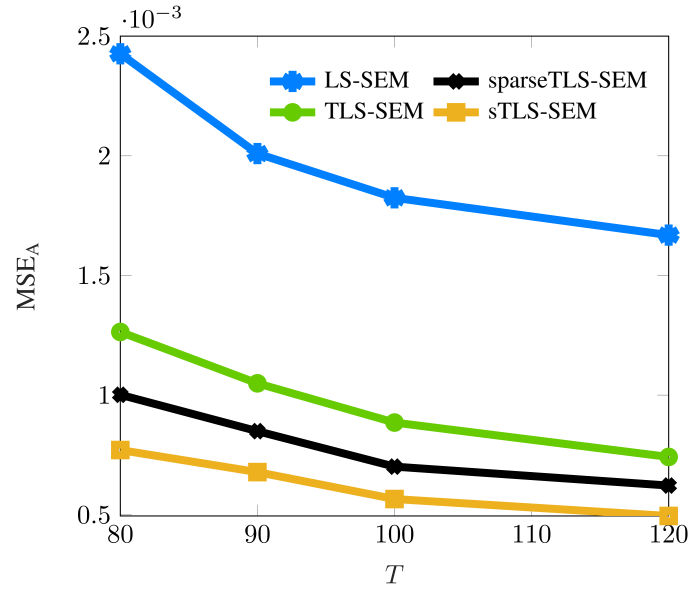

Figure 5: MSE_A as a function of T.

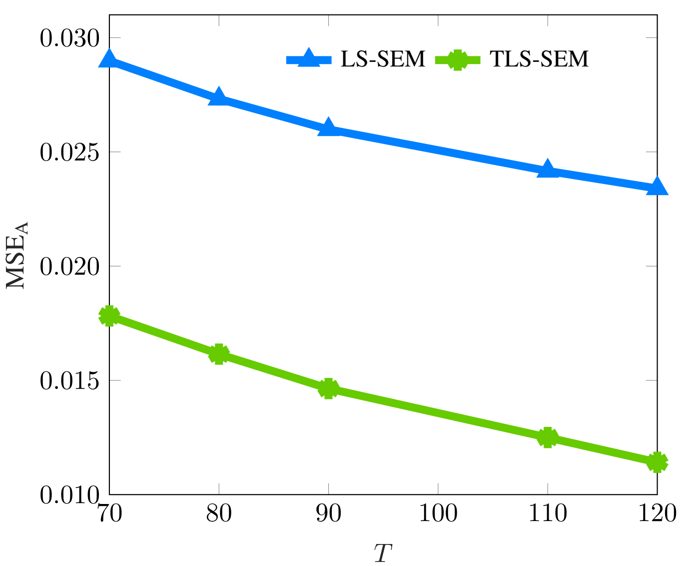

Figure 4: MSE_A versus T.

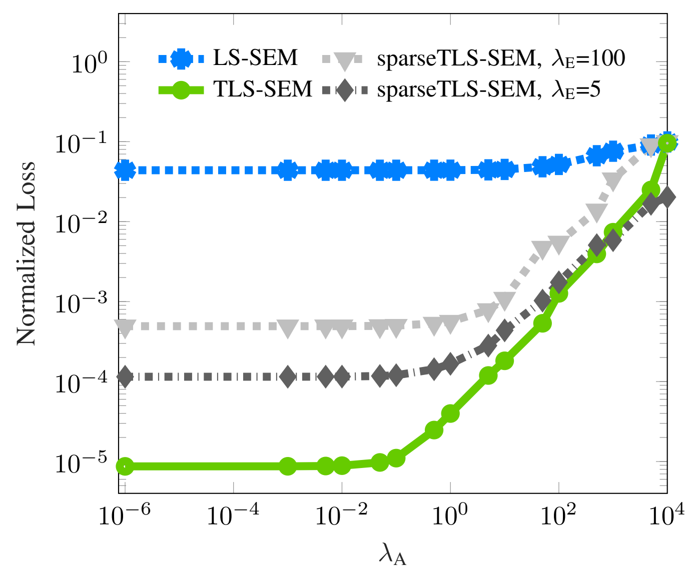

Figure 6: Loss function vs. λ_A for Gene regulatory network.

sively genotyped by the International HapMap project [22]. From the 929 identified genes, expression levels and the genotypes of the expression quantitative trait loci (eQTLs) of 39 immune-related genes were selected and normalized; see [11] and [46] for further details. Genotypes of eQTLs were adopted as known exogenous inputs $\mathbf{X}$, and gene expression levels were treated as the endogenous variables $\mathbf{Y}$. The underlying network as well as the matrix $\mathbf{B}$, were inferred by adopting TLS-SEM, sparseTLS, and LS-SEM methods.

Fig. 6 depicts the fitting loss divided by the norm of the data $\mathbf{Z}$, as $\|\mathbf{Z}-\mathbf{AZ}-\mathbf{BX}\|_F^2/\|\mathbf{Z}\|_F^2$ for LS-SEM, and $\|\mathbf{Z}-\mathbf{A}(\mathbf{Z}-\mathbf{E})-\mathbf{BX}\|_F^2/\|\mathbf{Z}\|_F^2$ for TLS-SEM. For all values of λ_A, i.e. the regularization parameter promoting the adjacency sparsity, TLS-SEM and sparseTLS-SEM outperform the LS-SEM, which implies that the inferred matrix $\mathbf{A}$ fits the model better when the signal perturbations are taken into account. When λ_A reaches very large values, all approaches perform similarly since the regularization term $\lambda_A\|\mathbf{A}\|_1$ prevails on all the other terms of the cost functions and $\hat{\mathbf{A}}$ becomes an all zero matrix. Furthermore, Fig. 7 illustrates the performance in terms of fitting error $\|\mathbf{Y}-\mathbf{AY}-\mathbf{BX}\|_F^2$, with $\mathbf{Y}=\mathbf{Z}-\mathbf{E}$ for TLS-SEM and sparseTLS, and $\mathbf{Y}=\mathbf{Z}$ for LS-SEM across values of λ_A. Again, TLS-SEM and sparseTLS-SEM outperform LS-SEM.

C. Signal inference under topology perturbations

We further tested the performance of the TLS algorithms in Sec. IV, and compared them with the conventional LS-SEM based signal recovery algorithm that does not account for topology perturbations. In this setting, the topology is perturbed and the goal is to identify $\mathbf{Y}$ from a subset of observations. A Kronecker graph with $N=27$ is generated as before.

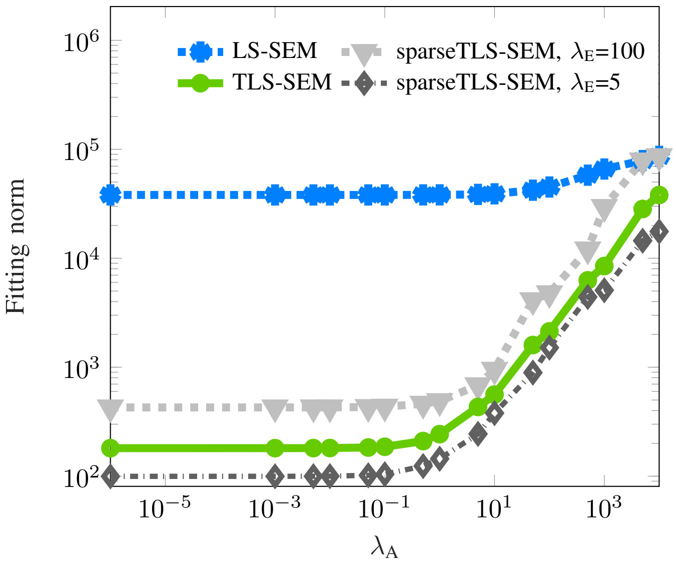

Figure 7: Fitting norm vs. λ_A for Gene regulatory network.

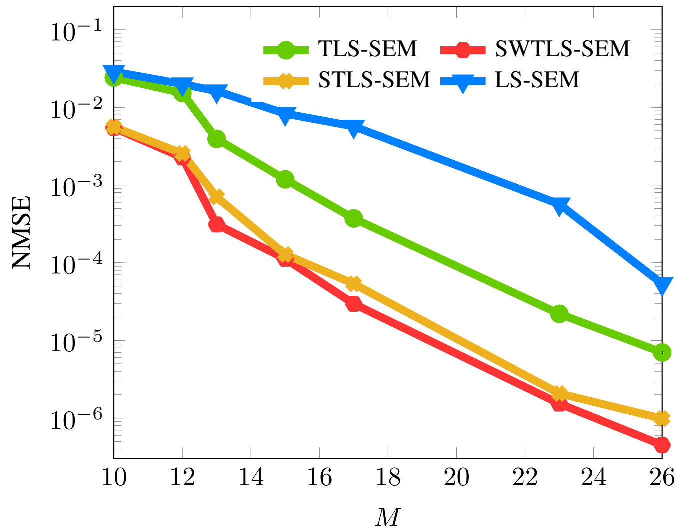

Figure 8: NMSE versus M.

With $T = 50$ and $\mathbf{B} = \mathbf{I}_N$, the entries are again randomly drawn as $[\mathbf{X}]_{ij} \sim \mathcal{U}[0, 3]$ and $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_i^2)$. Furthermore, we model the perturbation $\mathbf{\Delta}$ as Bernoulli$(\pi_l) \times [\mathbf{A}]_{ij}$, with $l := (v_i, v_j)$, which means that perturbations occur when one or more weighted links fail. In particular, $\pi_1 = \pi_2 = 0.9$, and $\pi_l \in [0.001, 0.02], l = 3, \ldots, L$, and we choose $r(\pi_l) = \frac{1}{\pi_l}$. Matrices $\mathbf{Y}$ and $\mathbf{\Psi}$ are then constructed according to (28) and (27), while λ_1 and λ_2 are selected via cross validation. Fig. 8 depicts the performance of LS-SEM, TLS-SEM, structured TLS under topology perturbations (STLS-SEM), and SWTLS-SEM in terms of normalized mean-square error

$$\text{NMSE} = \frac{\|\hat{\mathbf{Y}} - \mathbf{Y}\|_F^2}{\|\mathbf{Y}\|_F^2}. \tag{41}$$

The results are obtained by averaging over 1000 Monte Carlo realizations of $\mathbf{X}, \mathbf{E}, \mathbf{\Delta}$, and $\mathbf{D}_{\mathcal{S}}$. Fig. 8 shows the performance as a function of the number of sampled nodes M. Clearly, estimation performance improves as extra prior information is accounted for.

Figure 9 depicts the runtime (in seconds) of the topology perturbation aware TLS-SEM and SWTLS-SEM versus N, with $M = \lceil (2/3)N \rceil$. Each $\mathbf{A}$ is generated as Kronecker graph adjacency matrix of size N. As expected, the runtime increases with the number of nodes. However, the proposed TLS-SEM and SWTLS-SEM solvers are amenable to parallel implementation that would considerably reduce the runtime.

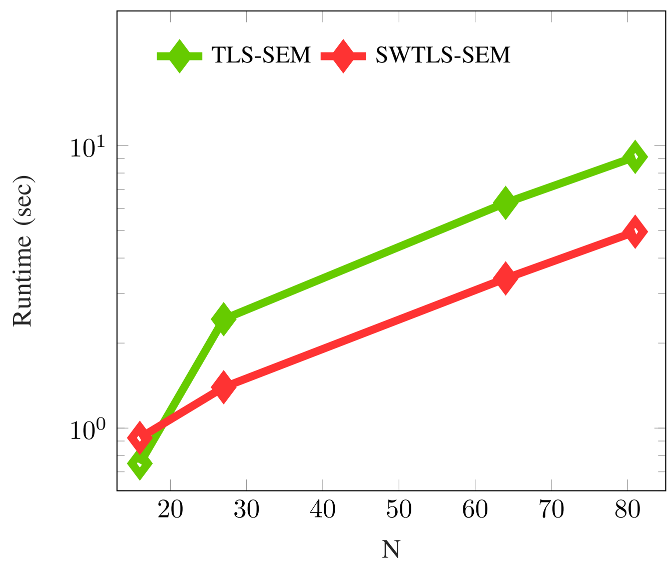

Figure 9: Runtime in seconds versus N.

D. Real tests for signal inference with topology perturbations

Finally, we test the proposed elTLS-based method in (30) to infer the signal given a subset of noisy observations and a perturbed graph topology.

The real data consists of path delay measurements on the Internet2 backbone [1]. The network has 9 nodes and 26 directed links. The delays are available for $N = 70$ paths per minute. Set $\{y_{nt}\}$ contains a subset of delays in milliseconds per path n and minute slot t. The known topologies are obtained based on the following three possible models.

M1. Here the paths connect origin-destination nodes by a series of links described by the path-link routing matrix $\Pi \in \{0, 1\}^{N \times 26}$, whose (n, l) entry is $\Pi_{n,l} = 1$ if path n traverses link l, and 0 otherwise. A graph is constructed with each vertex corresponding to one of these paths, and with the time-invariant adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ given by

$$\mathbf{A}_{n,n'} = \frac{\sum_{l=1}^{26} \Pi_{n,l} \Pi_{n',l}}{\sum_{l=1}^{26} \Pi_{n,l} + \sum_{l=1}^{26} \Pi_{n',l} - \sum_{l=1}^{26} \Pi_{n,l} \Pi_{n',l}} \tag{42}$$

for $n, n' = 1, \ldots, N$ and $n \neq n'$. The edge weight model in (42) assigns greater weights to edges connecting vertices whose associated paths share more links. This is reasonable because paths with common links usually experience similar delays [16].

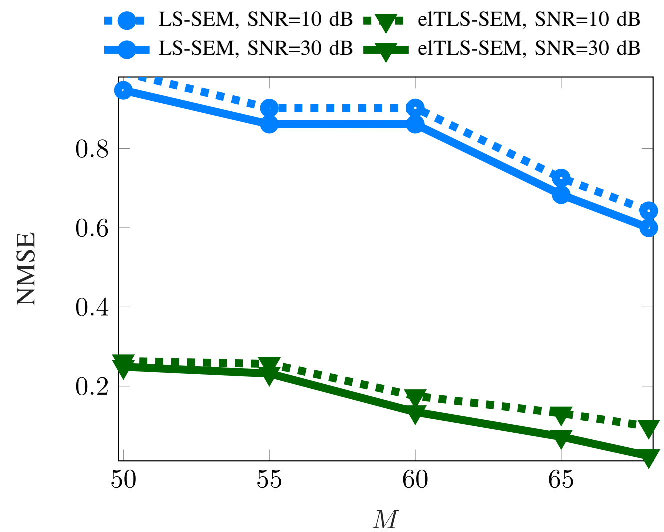

Figure 10: NMSE versus M, with $\mathbf{A}$ obtained via M1 and $T = 100$.

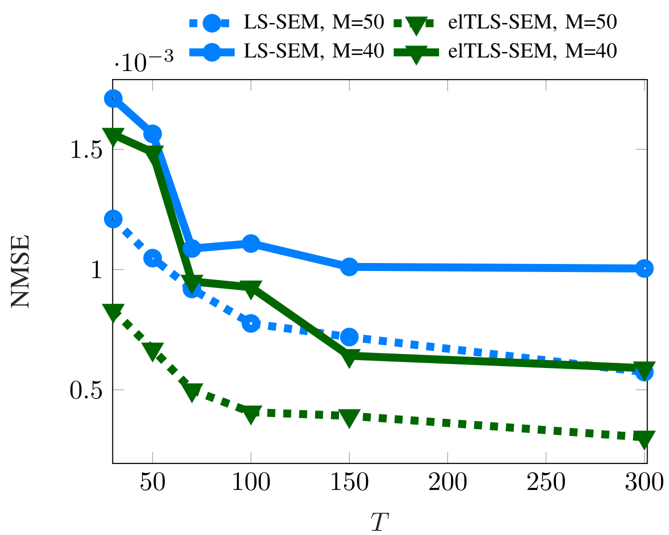

Figure 11: NMSE versus T, with $\mathbf{A}$ obtained via M2, SNR= 30 dB.

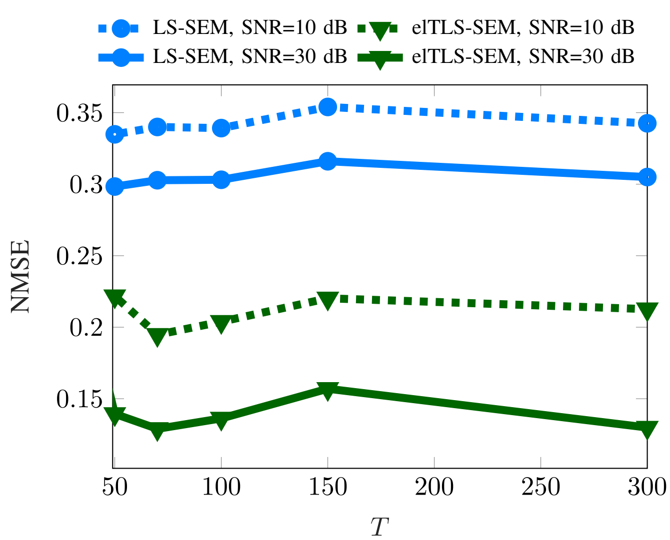

Figure 12: NMSE versus T, with $\mathbf{A}$ obtained M3 and $M = 50$.

M2. For the second topology, a training phase is introduced based on a subset of the signal observations, collected in the matrix $\mathbf{Y}_{train}$, to estimate the adjacency as the solution of

$$\min_{\mathbf{A}} \|\mathbf{Y}_{train} - \mathbf{A}\mathbf{Y}_{train}\|_F^2 \quad \text{s. to } a_{ii} = 0, \; i = 1, \ldots, N \tag{43}$$

where $\mathbf{Y}_{train} \in \mathbb{R}^{N \times T_{train}}$, with $T_{train} = 20$.

M3. The third topology is found as in (43), but the signals used for training are contaminated by noise, that is, $\bar{\mathbf{Y}}_{train} := \mathbf{Y}_{train} + \boldsymbol{\Xi}$, with $[\boldsymbol{\Xi}]_{ij} \sim \mathcal{N}(0, \sigma_\xi^2)$; while σ_ξ^2 is chosen such that $10 \log_{10}(\|\bar{\mathbf{y}}_{train}\|_F^2 / (N\sigma_\xi^2)) = -8$ dB, where $\bar{\mathbf{y}}_{train} \in \mathbb{R}^N$ is the average of the columns of $\mathbf{Y}_{train}$. Solving problem (43) with $\bar{\mathbf{Y}}_{train}$ instead of $\mathbf{Y}_{train}$ gives rise to an alternative topology with an inherent model mismatch. The observation error in (27) is generated using $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \sigma_{\boldsymbol{\varepsilon}}^2 \mathbf{I}), \forall t$.

Fig. 10 illustrates the NMSE versus the number of sampled nodes M when the topology is obtained from M1. It shows that the novel perturbation-aware elTLS-SEM outperforms the LS-SEM signal recovery approach by accounting for the possible model mismatch.

Figs. 11 and 12 illustrate the NMSE versus number of observations T with adjacency matrices obtained via M2 and M3, respectively. Once again, perturbation-aware elTLS-SEM outperforms LS-SEM signal recovery method. The performance gain of elTLS-SEM in Fig. 11 is less evident than that in Figures 10 and 12 because the adjacency matrix is obtained exactly following the SEM. Results are averaged over 100 realizations.

VI. Conclusions and Research Outlook

This contribution dealt with two challenging tasks over graphs, namely topology ID under signal perturbations, and signal inference under topology perturbation. To address the associated challenges, a spectrum of approaches based on total least-squares and structural equation models were developed. In addition, structured and weighted variants of TLS-SEM were introduced to flexibly account for extra prior information. Numerical tests on both synthetic and real data demonstrated the efficacy of the proposed algorithms.

Future research directions include distributed implementation of TLS-SEM to accommodate large-scale graphs, as well as generalizations of perturbed SEMs to account for nonlinear and dynamic inter-dependencies.

References

[1] “One-way ping internet2,” [Online], Available: http://software.internet2.
[2] I. G. Akrotirianakis and C. A. Floudas, “Computational experience with a new class of convex underestimators: Box-constrained nlp problems,” *J. Global Optimization*, vol. 29, no. 3, pp. 249–264, Jul. 2004.
[3] A. Anis, A. Gadde, and A. Ortega, “Efficient sampling set selection for bandlimited graph signals using graph spectral proxies,” *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3775–3789, Jul. 2016.
[4] M. Aoki and P. Yue, “On a priori error estimates of some identification methods,” *IEEE Trans. on Automatic Control*, vol. 15, no. 5, pp. 541–548, Oct. 1970.
[5] B. Baingana, G. Mateos, and G. B. Giannakis, “Proximal-gradient algorithms for tracking cascades over social networks,” *IEEE J. Sel Topics Sig. Process.*, vol. 8, no. 4, pp. 563–575, Aug. 2014.
[6] P. Balachandran, E. D. Kolaczyk, and W. D. Viles, “On the propagation of low-rate measurement error to subgraph counts in large networks,” *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 2025–2057, July 2017.

[7] J. A. Bazerque, B. Baingana, and G. B. Giannakis, "Identifiability of sparse structural equation models for directed and cyclic networks," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Austin, TX, 2013.

[8] A. Beck, A. Ben-tal, and M. Teboulle, "Finding a global optimal solution for a quadratically constrained fractional quadratic problem with applications to the regularized total least squares," *SIAM J. Matrix Anal. Appl*, vol. 28, no. 2, pp. 425–445, 2006.

[9] D. P. Bertsekas, "Nonlinear programming," *J. Oper. Res. Soc.*, vol. 48, no. 3, pp. 334–334, 1997.

[10] A. Bolstad, B. Van Veen, and R. Nowak, "Causal network inference via group sparse regularization," *IEEE Trans. Sig. Process.*, vol. 59, no. 6, pp. 2628–2641, June 2011.

[11] X. Cai, J. A. Bazerque, and G. B. Giannakis, "Inference of gene regulatory networks with sparse structural equation models exploiting genetic perturbations," *PLoS Comp. Biol.*, vol. 9, no. 5, p. Art. no. e1003068, May 2013.

[12] E. Ceci and S. Barbarossa, "Small perturbation analysis of network topologies," in *Proc. Intl. Conf. Acoust., Speech and Signal Process.* IEEE, Calgary, Canada, 2018, pp. 4194–4198.

[13] ——, "Robust graph signal processing in the presence of uncertainties on graph topology," in *Proc. of IEEE Workshop on Signal Process. Advances in Wireless Comm.*, Kalamata, Greece, 2018, pp. 1–5.

[14] E. Ceci, Y. Shen, G. B. Giannakis, and S. Barbarossa, "Signal and graph perturbations via total least-squares," in *Proc. of Asilomar Conf, on Signals, Systems, and Computers*, Pacific Grove, CA, 2018, pp. 747–751.

[15] J. Chang, E. D. Kolaczyk, and Q. Yao, "Estimation of edge density in noisy networks," *arXiv preprint arXiv:1803.02488*, Mar. 2018.

[16] D. B. Chua, E. D. Kolaczyk, and M. Crovella, "Network kriging," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 12, pp. 2263–2272, Dec. 2006.

[17] B. De Moor, "Structured total least-squares and L2 approximation problems," *Linear Algebra and its Applications*, vol. 188, no. 5, pp. 163–205, Aug. 1993.

[18] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society, series B*, vol. 39, no. 1, pp. 1–38, 1977.

[19] P. Di Lorenzo, S. Barbarossa, and P. Banelli, "Sampling and recovery of graph signals," in *Cooperative and Graph Signal Processing*, P. M. Djurić and C. Richard, Eds. Academic Press, 2018.

[20] P. Di Lorenzo and E. Ceci, "Online recovery of time-varying signals defined over dynamic graphs," in *Proc. of European Signal Process. Conf.*, Rome, Italy, 2018, pp. 131–135.

[21] W. Dinkelbach, "On nonlinear fractional programming," *Manag. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967.

[22] K. A. Frazer *et al.*, "A second generation human haplotype map of over 3.1 million SNPs," *Nature*, vol. 449, no. 7164, p. 851, 2007.

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

[24] G. B. Giannakis, Y. Shen, and G. V. Karanikolas, "Topology identification and learning over graphs: Accounting for nonlinearities and dynamics," *Proc. of the IEEE*, vol. 106, no. 5, pp. 787–807, 2018.

[25] A. S. Goldberger, "Structural equation methods in the social sciences," *Econometrica*, vol. 40, no. 6, pp. 979–1001, Nov. 1972.

[26] S. Haufe, R. Tomioka, G. Nolte, K. Müller, and M. Kawanabe, "Modeling sparse connectivity between underlying brain sources for eeg/meg," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, p. 1954–1963, May 2010.

[27] V. N. Ioannidis, M. Ma, A. N. Nikolakopoulos, G. B. Giannakis, and D. Romero, "Kernel-based inference of functions over graphs," in *Adaptive Learning Methods for Nonlinear System Modeling*. Elsevier, 2018, pp. 173–198.

[28] V. N. Ioannidis, D. Romero, and G. B. Giannakis, "Inference of spatio-temporal functions over graphs via multikernel kriged kalman filtering," *IEEE Trans. Signal Process.*, vol. 66, no. 12, pp. 3228–3239, 2018.

[29] V. N. Ioannidis, Y. Shen, and G. B. Giannakis, "Semi-blind inference of topologies and dynamical processes over dynamic graphs," *IEEE Trans. Sig. Process.*, vol. 67, no. 9, pp. 2263–2274, May 2019.

[30] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Filtering random graph processes over random time-varying graphs," *IEEE Trans. Sig. Process.*, vol. 65, no. 16, pp. 4406–4421, May 2017.

[31] E. P. Jiang and M. W. Berry, "Solving total least-squares problems in information retrieval," *Linear Algebra and its Applications*, vol. 316, no. 1-3, pp. 137–156, Jan 2000.

[32] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "An introduction to variational methods for graphical models," *Mach. Learn.*, vol. 37, no. 2, p. 183–233, Nov. 1999.

[33] A. Jung, "Learning the conditional independence structure of stationary time series: A multitask learning approach," *IEEE Trans. Signal Process.*, vol. 63, no. 21, pp. 5677–5690, Nov. 2015.

[34] A. Jung, G. Hannak, and N. Goertz, "Graphical lasso based model selection for time series," *IEEE Sig. Process. Letters*, vol. 22, no. 10, pp. 1781–1785, Oct. 2015.

[35] D. Kaplan, *Structural Equation Modeling: Foundations and Extensions*. Sage, 2009, vol. 2nd ed.

[36] M. Kolar and E. P. Xing, "On time varying undirected graphs," in *Proc. Intl. Conf. Art. Intell. Stat.*, 2011.

[37] G. Kollios, M. Potamias, and E. Terzi, "Clustering large probabilistic graphs," *IEEE Trans.on Knowl. and Data Engineering*, vol. 25, no. 2, pp. 325–336, Feb. 2013.

[38] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Appl.*, vol. 18, no. 2, pp. 95–138, 1977.

[39] S. L. Lauritzen, "The em algorithm for graphical association models with missing data," *Computational Statistics and Data Analysis*, vol. 19, no. 2, pp. 191 – 201, 1995.

[40] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani, "Kronecker graphs: An approach to modeling networks," *J. Mach. Learn. Res.*, vol. 11, pp. 985–1042, Mar. 2010.

[41] I. Markovsky and S. Van Huffel, "Overview of total least-squares methods," *Sig. Proc.*, vol. 87, no. 10, pp. 2283–2302, Apr. 2007.

[42] G. Mateos, S. Segarra, and A. Marques, "Inference of graph topology," in *Cooperative and Graph Signal Processing*, P. M. Djurić and C. Richard, Eds. Academic Press, 2018, pp. 349 – 374.

[43] B. Muthén, "A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators," *Psychometrika*, vol. 49, no. 1, pp. 115–132, Mar. 1984.

[44] S. K. Narang, A. Gadde, E. Sanou, and A. Ortega, "Localized iterative methods for interpolation in graph structured data," in *Proc. Global Conf. Signal Inf. Process.* IEEE, 2013, pp. 491–494.

[45] X. Ning and G. Karypis, "Slim: Sparse linear methods for top-n recommender systems," in *Proc. of Intl. Conf. on Data Mining*, Vancouver, Canada, 2011, pp. 497–506.

[46] J. K. Pickrell *et al.*, "Understanding mechanisms underlying human gene expression variation with RNA sequencing," *Nature*, vol. 464, no. 7289, p. 768, 2010.

[47] N. T. Quang and A. Jung, "Learning conditional independence structure for high-dimensional uncorrelated vector processes," in *Proc. Intl. Conf. Acoust., Speech and Signal Process.*, 2017, pp. 5920–5924.

[48] Y. Shen, B. Baingana, and G. B. Giannakis, "Tensor decompositions for identifying directed graph topologies and tracking dynamic networks," *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3675–3687, July 2017.

[49] A. J. Smola and R. I. Kondor, *Kernels and regularization on graphs*, ser. Learning Theory and Kernel Machines. New York: Springer, 2003.

[50] J. Songsiri, J. Dahl, and L. Vanderberghe, "Maximum-likelihood estimation of autoregressive models with conditional independence contraints," in *Proc. Intl. Conf. Acoust., Speech and Signal Process.*, 2009, pp. 1701–1704.

[51] P. Sprent, "Models in regression and related topics," *Methuen's Monographs on Applied Probability and Statistics*, Oct. 1969.

[52] P. A. Traganitis, Y. Shen, and G. B. Giannakis, "Network topology inference via elastic net structural equation models," in *European Signal Process. Conf.*, Kos Island, Greece, Sep. 2017, pp. 146–150.

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

[54] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, ser. Frontier in Applied Mathematics. Philadelphia, PA: SIAM, 1991, vol. 9.

[55] G. Wang, V. Kekatos, A. J. Conejo, and G. B. Giannakis, "Ergodic energy management leveraging resource variability in distribution grids," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4765–4775, Nov. 2016.

[56] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*. New York, NY, USA: Oxford University Press, Inc., 1988.

[57] H. Zhu, A. Cano, and G. B. Giannakis, "Distributed consensus-based demodulation: Algorithms and error analysis," *IEEE Trans. on Wireless Communications*, vol. 9, no. 6, pp. 2044–2054, June 2010.

[58] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2002–2016, May 2011.

[59] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. of the Royal Stat. Soc.: Series B*, vol. 67, no. 2, pp. 301–320, 2005.