Signal and Graph Perturbations via Total Least-Squares

Elena Ceci[1], Yanning Shen[2], Georgios B. Giannakis[2], Sergio Barbarossa[1]
[1]Sapienza University of Rome, DIET, Via Eudossiana 18, 00184, Rome, Italy
[2]Dept. of ECE and DTC, University of Minnesota, Minneapolis, USA
Emails: elena.ceci@uniroma1.it, shenx513@umn.edu, georgios@umn.edu, sergio.barbarossa@uniroma1.it

***Abstract*—Graphs are pervasive in various applications capturing the complex behavior observed in biological, financial, and social networks, to name a few. Two major learning tasks over graphs are topology identification and inference of signals evolving over graphs. Existing approaches typically aim at identifying the topology when signals on all nodes are observed, or, recovering graph signals over networks with known topologies. In practice however, signal or graph perturbations can be present in both tasks, due to model mismatch, outliers, outages or adversaries. To cope with these perturbations, this work introduces regularized total least-squares (TLS) based approaches and corresponding alternating minimization algorithms with convergence guarantees. Tests on simulated 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 are widely adopted for analyzing interactions among nodes in e.g. biological or financial systems, where data-driven networks are constructed to capture (un)directed dependencies. They are also useful for modeling and understanding properties of physical networks. For example, analyzing vehicular, power, or communication networks is crucial for resource allocation tasks. However, perturbations on links or vertices may be present in both data-driven and physical networks. In data-driven networks for instance, links of the inferred topology may be uncertain due to e.g., model mismatch. While in physical networks, topologies may be perturbed due to e.g., links outages, and the signals observed over nodes may deviate from their real values because of outliers or node outages.

Perturbation analysis over graphs has been widely studied recently. Error propagation analysis has been carried out with respect to network characteristics, such as subgraphs counts [3], [8]. Other works treat perturbations through probabilistic or uncertain graphs, in the context of clustering [15], graph filtering [13], or consensus [23]. Topology perturbations have been also considered for robust resource allocation and graph signal inference over uncertain networks [6], [7]; see further [10] for tracking graph signals over dynamic undirected networks, and [6], [7] for using small perturbation analysis of the Laplacian matrix [22].

Structural equation models (SEMs) have well-documented merits for network topology inference thanks to their simplicity and ability to capture directed dependencies among nodes [2], [5], [11], [14]. Prior works on SEMs mostly assume that signals over all nodes are observed. Leveraging piecewise stationarity, SEM-based topology inference was pursued in [18] when just (partial) statistics of nodal measurements are given, while a joint inference algorithm was developed in [12] to identify the topology as well as interpolating graph signals when only partial observations of the nodal signals are available. However, neither [18] nor [12] account for perturbations on the observations.

In this work, signal and graph perturbations are considered for topology identification or graph signal inference tasks carried using total least-squares (TLS). TLS is a widely-used tool for analyzing perturbed linear systems that is also known under the term 'error-in-variables' in statistical learning [19]. TLS is a generalization of least-squares (LS) for fully-perturbed linear models that allow error mismatch (a.k.a. noise) to be present in both the input and the output matrices [21]. TLS and regularized TLS emerge in several application fields, including system identification, information retrieval, and reconstruction of medical images [17]. Building upon TLS, weighted TLS [1], structured TLS [9], and sparse TLS [24] have been introduced to incorporate different prior information. Based on these prior works, two pertinent graph learning tasks with well-appreciated applications will be investigated: Identifying the topology from perturbed nodal measurements; and, inferring nodal signals over perturbed graphs based on TLS and SEMs. In other words, possible perturbations are accounted for, either in the nodal measurements, or, in the graph structure for SEMs.

Notation. Bold lower (upper) case letters denote column vector, e.g., $\boldsymbol{a}$ (matrix $\mathbf{A}$), while operators $(\cdot)^\top$, and $\mathrm{vec}(\cdot)$ stand for transposition, and column-wise matrix vectorization, respectively. A $K \times K$ identity matrix is denoted by $\mathbf{I}_K$, and $\boldsymbol{e}_i$ stands for the i-th canonical vector; while $\mathrm{diag}(\boldsymbol{a})$ represents a diagonal matrix with entries of $\boldsymbol{a}$ on its diagonal. Finally, the ℓ_1, ℓ_2, and Frobenius norms will be denoted by $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_F$, respectively.

II. Preliminaries

In this section, we will briefly introduce SEMs and TLS, along with structured and weighted TLS variants.

A. Structural Equation Models

Consider an N-node directed network with unknown adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$, and let y_{it} denote the t-th

Work in this paper was supported by grants NSF 1711471, 1500713 and NIH 1R01GM104975-01.

measurement at node i. SEMs postulate that y_{it} depends on its single-hop neighbors, and an exogenous input x_{it}; that is,

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

where $a_{ij} := [\mathbf{A}]_{ij}$, and $a_{ij} \neq 0$ signifies that a directed edge from j to i is present. Concatenating nodal measurements into $\boldsymbol{y}_t := [y_{1t}, \ldots, y_{Nt}]^\top$, and $\boldsymbol{x}_t := [x_{1t}, \ldots, x_{Nt}]^\top$ per slot t, the matrix-vector version of (1) can be compactly written as $\boldsymbol{y}_t = \mathbf{A}\boldsymbol{y}_t + \mathbf{B}\boldsymbol{x}_t,\ t = 1 \ldots T$, where $[\mathbf{A}]_{ii} = 0$ and $\mathbf{B} := \text{diag}(b_{11}, \ldots, b_{NN})$. Collecting observations over T instants in a matrix $\mathbf{Y} := [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T]$, yields the noiseless matrix SEM

$$\mathbf{Y} = \mathbf{AY} + \mathbf{BX}\,. \tag{2}$$

With $\mathbf{Y}$ and $\mathbf{X}$ available, prior works on SEMs develop algorithms to obtain $\mathbf{B}$ and $\mathbf{A}$. To focus on modeling and analyzing perturbations on $\mathbf{A}$ and $\mathbf{Y}$, we will consider here that $\mathbf{B}$ has been found from historical data.

In the presence of noise, prior works on SEMs simply consider perturbations as additive observation noise; that is, $\mathbf{Y} = \mathbf{AY} + \mathbf{BX} + \mathbf{V}$, where $\mathbf{V} \in \mathbb{R}^{N \times T}$ denotes the error matrix. Given measurements $\mathbf{Y}$, $\mathbf{X}$ (and here $\mathbf{B}$ too), the adjacency matrix $\mathbf{A}$ can be estimated via LS or regularized LS [4], [5]. On the other hand, given $\mathbf{A}$, $\mathbf{BX}$ and a subset of entries of $\mathbf{Y}$, it is also possible to interpolate the unobserved nodal signals using LS-based methods [12]. However, such a noisy model does not capture possible perturbations in $\mathbf{A}$, and as will be shown later, LS is not sufficient to account for the error of SEMs, which is by its nature, a self-dictionary model. This motivates our adoption of TLS to cope with graph signal and topology perturbations present in SEMs.

B. Weighted and structured TLS

Consider the general *under-determined* linear system of equations $\boldsymbol{\Phi} = \boldsymbol{H}\boldsymbol{\Theta} + \mathbf{V}$, where $\boldsymbol{\Phi} \in \mathbb{R}^{M \times T}$ denotes the output matrix with $M < T$, $\boldsymbol{H} \in \mathbb{R}^{M \times N}$ the input (or regression) matrix, $\boldsymbol{\Theta} \in \mathbb{R}^{N \times M}$ the unknown matrix, and $\mathbf{V} \in \mathbb{R}^{M \times T}$ the additive noise matrix. Different from classical LS where one considers output noise only, TLS treats symmetrically input and output in the sense that both $\boldsymbol{H}$ and $\boldsymbol{\Phi}$ can have errors that arise due to model mismatch, noise, and outliers that can be present in various applications. Accounting for such errors, the linear model becomes $\boldsymbol{\Phi} = (\boldsymbol{H} + \boldsymbol{P})\boldsymbol{\Theta} + \mathbf{V}$, where $\boldsymbol{P}$ denotes the input noise matrix.

Based on the latter, TLS solves the following problem

$$\begin{aligned} &\min_{\boldsymbol{\Theta}, \boldsymbol{P}} \|[\boldsymbol{P}, \mathbf{V}]\|_F^2 \\ &\text{s. to} \qquad \boldsymbol{\Phi} = (\boldsymbol{H} + \boldsymbol{P})\boldsymbol{\Theta} + \mathbf{V}. \end{aligned} \tag{3}$$

Whenever possible, exploiting the input and output structure, as well as prior information available about noise statistics, is expected to yield improved performance of TLS estimates. In our context, structure refers to the following [17], [24].

Definition 1. Given a parameter vector $\boldsymbol{\omega} \in \mathbb{R}^{n_\omega}$, the $M \times (N + T)$ data matrix $[\boldsymbol{H} \quad \boldsymbol{\Phi}](\boldsymbol{\omega})$ has a structure $\mathscr{S}(\boldsymbol{\omega})$ characterized by $\boldsymbol{\omega}$, if and only if there is a mapping such that $\boldsymbol{\omega} \in \mathbb{R}^{n_\omega} \rightarrow [\boldsymbol{H} \quad \boldsymbol{\Phi}](\boldsymbol{\omega}) := \mathscr{S}(\boldsymbol{\omega}) \in \mathbb{R}^{M \times (N+T)}$.
We can take advantage of the structure when $n_\omega \ll M(N+T)$, because then $\boldsymbol{\omega}$ provides a parsimonious representation of the data matrix. Examples of structured matrices include Toeplitz and Hankel ones that are present in system identification, deconvolution, and linear prediction, or Vandermonde and circulant matrices that show up in e.g., spatio-temporal harmonic retrieval problems [17]. Note that Definition 1 reduces to the trivial case $\boldsymbol{\omega} := \text{vec}([\boldsymbol{H} \quad \boldsymbol{\Phi}])$ with dimension $M(N + T)$, which corresponds to the unstructured case. Consider now introducing the parameter vector $\boldsymbol{\omega}$ and the noise vector $\boldsymbol{\nu} \in \mathbb{R}^{n_\omega}$, such that $\mathscr{S}(\boldsymbol{\omega} + \boldsymbol{\nu}) := [\boldsymbol{H} + \boldsymbol{P} \quad \boldsymbol{\Phi} + \mathbf{V}](\boldsymbol{\omega} + \boldsymbol{\nu})$. The Frobenious norm in (3) becomes the ℓ_2 norm of $\boldsymbol{\nu}$.

The weighted TLS is obtained if prior knowledge on the errors is incorporated by weighting the norm in (3) through matrix $\mathbf{W}$. Jointly, the structured and weighted (SW) version of the original TLS cost $\|\boldsymbol{P} \quad \boldsymbol{V}\|_F^2$ is expressed as $\boldsymbol{\nu}^\top \mathbf{W} \boldsymbol{\nu}$, where $\mathbf{W} \succeq \mathbf{0} \in \mathbb{R}^{n_\omega \times n_\omega}$. Clearly, with $\mathbf{W} = \mathbf{I}$, the SWTLS cost reduces to a structured-only form.

III. Topology ID with signal perturbations.

In both physical and data-driven networks, nodal signals may be perturbed due to outliers or defects in the measuring process. In this case, consider the topology identification (ID) problem with noisy nodal observations $\mathbf{Z} = \mathbf{Y} + \mathbf{E}$ available. Substituting $\mathbf{Y} = \mathbf{Z} - \mathbf{E}$ into (2), yields the "measurement-perturbed" SEM as

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

where a common error is present both in the output (O) and input (I) matrices. Equation (3) shows how the TLS-based approach deals with I/O errors.

Because most real-world networks are not densely connected, it is reasonable to consider that the adjacency matrix $\mathbf{A}$ is sparse, which is the case with e.g., social, transportation, and biological networks. Accounting for the latter through a sparsity-promoting regularization term, (4) boils down to the regularized TLS-based approach to SEM (TLS-SEM) given by

$$\{\hat{\mathbf{A}}, \hat{\mathbf{E}}\} = \arg\min_{\mathbf{A}, \mathbf{E}} \|[\mathbf{E}, \mathbf{E}]\|_F^2 + \lambda \|\mathbf{A}\|_1 \tag{5a}$$

$$\text{s.to} \quad \mathbf{Z} = \mathbf{A}(\mathbf{Z} - \mathbf{E}) + \mathbf{BX} + \mathbf{E} \tag{5b}$$

$$[\mathbf{A}]_{ii} = 0, \ \ i = 1, \ldots, N \tag{5c}$$

where λ is a non-negative regularization parameter, and constraint (5c) enforces the absence of self-loops in $\mathbf{A}$.

Substituting constraint (5b) in one of the errors in (5a), leads to the equivalent formulation

$$\begin{aligned} \{\hat{\mathbf{A}}, \hat{\mathbf{E}}\} = &\arg\min_{\mathbf{A}, \mathbf{E}} \|\mathbf{Z} - \mathbf{A}(\mathbf{Z} - \mathbf{E}) - \mathbf{BX}\|_F^2 \\ &+ \|\mathbf{E}\|_F^2 + \frac{\lambda}{2} \|\mathbf{A}\|_1 \\ &\text{s.to} \qquad [\mathbf{A}]_{ii} = 0, \quad i = 1, \ldots, N. \end{aligned} \tag{6}$$

Problem in (6) is per-block convex wrt each block variable. Given $\hat{\mathbf{A}}[k]$, we estimate the error in iteration $k+1$ as

$$\hat{\mathbf{E}}[k+1] = \arg\min_{\mathbf{E}} \left\|(\mathbf{Z} - \hat{\mathbf{A}}[k]\mathbf{Z}) + \hat{\mathbf{A}}[k]\mathbf{E} - \mathbf{B}\mathbf{X}\right\|_F^2 + \|\mathbf{E}\|_F^2 \tag{7}$$

which admits a closed-form solution

$$\hat{\mathbf{E}}[k+1] = (\hat{\mathbf{A}}^\top[k]\hat{\mathbf{A}}[k] + \mathbf{I}_N)^{-1}\hat{\mathbf{A}}^\top[k](\hat{\mathbf{A}}[k]\mathbf{Z} + \mathbf{B}\mathbf{X} - \mathbf{Z}) \tag{8}$$

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

$$\hat{\mathbf{A}}[k+1] = \arg\min_{\mathbf{A}} \left\|(\mathbf{Z} - \mathbf{A}(\mathbf{Z} - \hat{\mathbf{E}}[k+1]) - \mathbf{B}\mathbf{X}\right\|_F^2 + \frac{\lambda}{2}\|\mathbf{A}\|_1 \tag{9}$$

As the objective in (9) is nonsmooth but strongly convex, it admits a globally optimal solution. A proximal gradient iteration can be employed to cope with the non-smooth regularizer. The derivation of the algorithm is omitted due to the lack of space, but can be deduced from [2]. Notice that problem (6) is still per-block convex when $\mathbf{B}$ is also a variable to be estimated [2]. The alternating minimization method under regularity conditions is guaranteed to converge at least to a stationary point, as asserted in the following proposition.

Proposition 1. *The iterates* $\{\hat{\mathbf{E}}[k], \hat{\mathbf{A}}[k]\}$ *resulting from the minimization of* (7) *and* (9) *converge monotonically at least to a stationary point of problem* (6).

Proof. See [20]. □

IV. Signal inference with topology perturbations

In several applications such as communications and power networks, edges may drop due to link failures. Likewise for data-driven networks, errors in the measurement collection process, and model mismatch effects suggest modeling links as being uncertain. In this context, the goal of this section is graph signal reconstruction, when only a subset of nodal measurements denoted by $\mathcal{S}$ is given, along with the perturbed graph topology. The observation model can then be written as $\boldsymbol{\psi}_t = \mathbf{D}_{\mathcal{S}}\boldsymbol{y}_t + \boldsymbol{\varepsilon}_t$ for $t = 1, \ldots, T$, where $\boldsymbol{\psi}_t \in \mathbb{R}^M$, and M denotes the cardinality of $\mathcal{S}$. Let $\mathbf{D}_{\mathcal{S}} \in \mathbb{R}^{M\times N}$ denote the selection matrix formed to have rows (with indices in the set $\mathcal{S}$) of the $N \times N$ identity matrix, and $\boldsymbol{\varepsilon}_t \in \mathbb{R}^M$ is the observation error vector at slot t. Stacking T observations to form $\boldsymbol{\Psi} := [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_T]$, and likewise for $\boldsymbol{\mathcal{E}} := [\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_T]$, the matrix model boils down to

$$\boldsymbol{\Psi} = \mathbf{D}_{\mathcal{S}}\mathbf{Y} + \boldsymbol{\mathcal{E}}. \tag{10}$$

Let now $\mathbf{A}_0$ denote the known and possibly perturbed nominal adjacency matrix, $\boldsymbol{\Delta} \in \mathbb{R}^{N\times N}$ the topology perturbation matrix. The linear SEM in (2) then reduces to

$$\mathbf{Y} = (\mathbf{A}_0 - \boldsymbol{\Delta})\mathbf{Y} + \mathbf{B}\mathbf{X} \tag{11}$$

where $\mathbf{A}_0 - \boldsymbol{\Delta}$ is the unknown adjacency matrix.

With $\mathbf{B}\mathbf{X}$ acquired as mentioned in the previous section, and bearing in mind that TLS can account for both input and output errors (here $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{\Delta}$), we formulate our regularized TLS-SEM task as (cf. (10) and (11))

$$\{\hat{\boldsymbol{\Delta}}, \hat{\boldsymbol{\mathcal{E}}}, \hat{\mathbf{Y}}\} = \arg\min_{\boldsymbol{\Delta},\boldsymbol{\mathcal{E}},\mathbf{Y}} \lambda_1||[\boldsymbol{\Delta}||_1 + \lambda_2||\boldsymbol{\mathcal{E}}]||_F^2 + ||\mathbf{Y} - (\mathbf{A}_0 - \boldsymbol{\Delta})\mathbf{Y} - \mathbf{B}\mathbf{X}||_F^2 \tag{12}$$

$$\text{s. to} \quad \boldsymbol{\Psi} = \mathbf{D}_{\mathcal{S}}\mathbf{Y} + \boldsymbol{\mathcal{E}} \tag{13}$$

$$[\boldsymbol{\Delta}]_{ii} = 0, \forall i \tag{14}$$

where the ℓ_1-norm promotes sparsity of the perturbed links, while the fitting term takes into account the perturbed SEM that allows us to reconstruct the graph signal over the entire network. Upon substituting the constraint (13) into the cost function, we arrive at

$$\{\hat{\boldsymbol{\Delta}}, \hat{\mathbf{Y}}\} = \arg\min_{\boldsymbol{\Delta},\mathbf{Y}} \lambda_1\|\boldsymbol{\Delta}\|_1 + \lambda_2\|\boldsymbol{\Psi} - \mathbf{D}_{\mathcal{S}}\mathbf{Y}\|_F^2 + \|\mathbf{Y} - (\mathbf{A}_0 - \boldsymbol{\Delta})\mathbf{Y} - \mathbf{B}\mathbf{X}\|_F^2 \quad \text{s.to} \quad [\boldsymbol{\Delta}]_{ii} = 0, \forall i\ . \tag{15}$$

Problem (15) is per-block convex, and thus it can be solved iteratively via alternating minimization with guaranteed convergence, to at least a stationary point; see [20] and Prop. 1.

A. Structured and weighted TLS for topological perturbations

In this subsection, we will leverage the structure of the nominal adjacency matrix along with prior information on the errors to formulate a structured and weighted TLS problem (cf. Sec. II-B). If the nominal network has L links, the nominal adjacency matrix has the following structure (cf. Definition 1)

$$\mathbf{A}_0 = \mathscr{S}(\boldsymbol{\omega}) := \sum_{l=1}^{L} \omega_l \mathbf{S}_l^{\mathbf{A}} \tag{16}$$

where $\mathbf{S}_l^{\mathbf{A}} := \mathbf{i}_l \cdot \mathfrak{f}_l^\top$ with $\mathbf{i}_l$ denoting the $N \times 1$ vector having all-zero entries except one entry that equals 1 at the position of the source node of link l, and $\mathfrak{f}_l$ the $N \times 1$ vector of all-zero entries except one that equals 1 at the position of the sink node of link l. Let now $\boldsymbol{\omega} := [\omega_1, \ldots \omega_L]^\top$ be the vector collecting edge weights characterizing the structure $\mathscr{S}(\boldsymbol{\omega})$ of $\mathbf{A}_0$. Such a structure accounts for the (non)zero entries of the adjacency matrix, and gives rise to a form having reduced the number of unknowns from N^2 to L.

In certain application domains, additional information can be given on the failure probabilities π_l for $l = 1, \ldots, L$, while the noise variance σ_i^2 can be also known across nodes $i = 1, \ldots, N$. Such prior information can be available after observing a network over time, and collecting the occurrence of failures, along with statistics of the measurement errors.

First, we will recast here the TLS formulation to account for the structure of the nominal adjacency matrix. Based on Sec. II-B and (16), the parameter vector of $\mathbf{A}_0$ is $\boldsymbol{\omega}$, while that for the error is $\boldsymbol{\nu}_{\boldsymbol{A}} := [\nu_1^A, \ldots, \nu_L^A]^\top$ with entries being nonzero when a failure or an edge weight alteration occurs. Likewise for the SWTLS cost in Section II-B, the first two terms in (12) become the weighted ℓ_1-norm of the topology error vector $\|\mathbf{W}_A\boldsymbol{\nu}_{\boldsymbol{A}}\|_1$, and the weighted Frobenious norm of the observation error $\|\boldsymbol{\mathcal{E}}\|_{\mathbf{W}_\Psi}^2$, where $\mathbf{W}_A$ and $\mathbf{W}_\Psi$ are respectively the weight matrices of the topology and the measurement errors.

Having introduced the structure and the weights, we will now formulate the unconstrained regularized structured and weighted (SW) TLS problem based on SEM, as follows

$$\min_{\mathbf{Y},\boldsymbol{\nu}^{\mathbf{A}},\boldsymbol{\mathcal{E}}} \lambda_1 \left\|\mathbf{W}_A \boldsymbol{\nu}^{\mathbf{A}}\right\|_1 + \lambda_2 \left\|\boldsymbol{\Psi} - \mathbf{D}_{\mathcal{S}}\mathbf{Y}\right\|^2_{\mathbf{W}_\Psi} + \left\|\mathbf{Y} - \sum_{l=1}^{L}(\omega_l + \nu_l^{\mathbf{A}})\mathbf{S}_l^{\mathbf{A}}\mathbf{Y} - \mathbf{B}\mathbf{X}\right\|_F^2 \quad (17)$$

where matrix $\mathbf{W}_A := [\text{diag}(\pi_1 \dots \pi_L)]^{-1}$, and $\mathbf{W}_\Psi := [\text{diag}(\sigma_1^2 \dots \sigma_M^2)]^{-1}$. Problem (17) is per-block convex, and can be solved again via alternating minimization. Given $\hat{\boldsymbol{\nu}}_{\mathbf{A}}[k]$ at iteration k, the graph signal at $k+1$ is reconstructed as

$$\hat{\mathbf{Y}}[k+1] = \arg\min_{\mathbf{Y}} \lambda_2 \left\|\boldsymbol{\Psi} - \mathbf{D}_{\mathcal{S}}\mathbf{Y}\right\|^2_{\mathbf{W}_\Psi} + \left\|\mathbf{Y} - \sum_{l=1}^{L}(\omega_l - \hat{\nu}_l^{\mathbf{A}}[k])\mathbf{S}_l^{\mathbf{A}}\mathbf{Y} - \mathbf{B}\mathbf{X}\right\|_F^2. \quad (18)$$

This sub-problem is also convex, and admits the closed-form update

$$\hat{\mathbf{Y}}[k+1] = (\mathbf{C}^\top[k]\mathbf{C}[k] + \lambda_2\mathbf{D}_{\mathcal{S}}^\top\mathbf{W}_\Psi\mathbf{D}_{\mathcal{S}})^{-1}(\mathbf{C}^\top[k]\mathbf{B}\mathbf{X} + \lambda_2\mathbf{D}_{\mathcal{S}}^\top\mathbf{W}_\Psi\boldsymbol{\Psi}) \quad (19)$$

where $\mathbf{C}[k] := (\mathbf{I} - \sum_{l=1}^{L}(\omega_l + \hat{\nu}_l^{\mathbf{A}}[k])\mathbf{S}_l^{\mathbf{A}})$.

Given $\hat{\mathbf{Y}}[k+1]$, the estimate $\hat{\boldsymbol{\nu}}_A[k+1]$ is obtained as

$$\hat{\boldsymbol{\nu}}_A[k+1] = \arg\min_{\boldsymbol{\nu}_{\mathbf{A}}} \lambda_1 \left\|\mathbf{W}_A\boldsymbol{\nu}_{\mathbf{A}}\right\|_1 + \left\|\hat{\mathbf{Y}}[k+1] - \sum_{l=1}^{L}(\omega_l - \nu_l^{\mathbf{A}})\mathbf{S}_l^{\mathbf{A}}\hat{\mathbf{Y}}[k+1] - \mathbf{B}\mathbf{X}\right\|_F^2. \quad (20)$$

Sub-problem (20) is again convex, but not differentiable. For this reason, we employ an iterative proximal gradient solver with provably guaranteed convergence to at least a stationary point, as asserted by [20] and Prop. 1.

V. Numerical Tests

Topology identification with signal perturbations. Here, we test the performance of the iterative solver of (6), and compare it with the conventional LS method based on SEM [2], [5]. The goal is to identify $\mathbf{A}$ when the signal measurements are perturbed. A Kronecker graph with $N = 64$ is generated as in [16]. With $T = 120$, the entries of $\mathbf{X}$ are drawn i.i.d. from the uniform distribution $\mathcal{U}[0, 1.5]$, and those of $\mathbf{E}$ from the Gaussian $\mathcal{N}(0, \sigma_E^2)$. Matrices $\mathbf{Y}$ and $\mathbf{Z}$ are then constructed according to (2) and (4), while λ is selected via cross-validation. Figure 1 depicts the performance of LS-SEM and TLS-SEM in terms of mean-square error (MSE), defined as $\text{MSE} := \sum_{ij}(\hat{a}_{ij} - a_{ij})^2/N^2$. In this figure, the two methods are compared as a function of σ_E. We observe that for small values of σ_E, the methods perform comparably, but when σ_E increases the TLS-SEM outperforms the SEM approach, until the point when the error becomes too large. Figure 2 compares

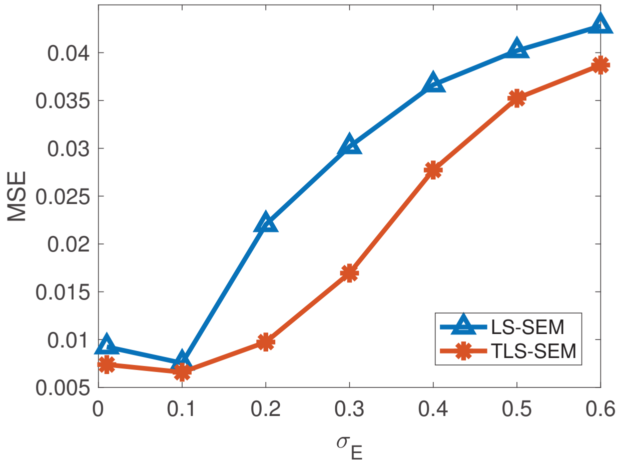

Figure 1: MSE versus Error standard deviation.

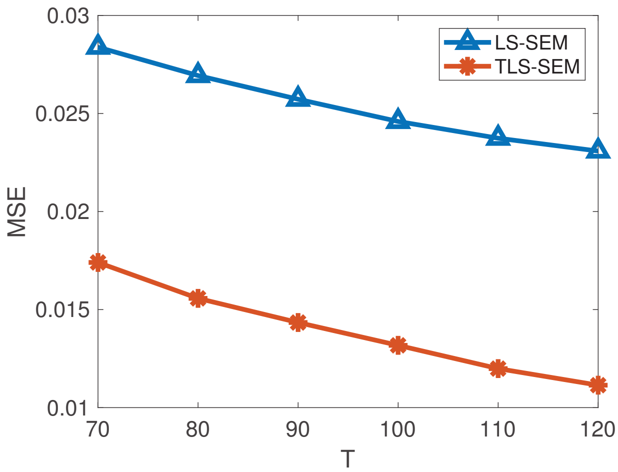

Figure 2: MSE versus number of observations.

LS-SEM with TLS-SEM in terms of MSE as a function of the number of observations T, with the standard deviation of errors fixed to $\sigma_E = 0.2$.

These numerical results suggest that TLS-SEM outperforms LS-SEM even when the number of observations is small.

Signal inference with topology perturbations. We further tested the performance of the algorithm in Section IV, and compared it with the conventional LS-SEM. Here, 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. With $T = 50$ and $\mathbf{B} = \mathbf{I}$, the entries of $\mathbf{X}$ and $\mathbf{E}$ are drawn i.i.d. from uniform $\mathcal{U}[0, 3]$, and Gaussian $\mathcal{N}(0, \sigma_{n_{ij}}^2)$ distributions, respectively. Furthermore, we used $Bernoulli(P_l) \times [\mathbf{A}]$ to model the perturbation $\boldsymbol{\Delta}$, meaning that perturbations occur when one or more weighted links fail. In particular, $P_1 = P_2 = 0.7$, and $P_l \in [0.001, 0.02], l = 3, \dots, L$. Matrices $\mathbf{Y}$ and $\boldsymbol{\Psi}$ are then constructed according to (11) and (10), while λ_1 and λ_2 are selected via cross validation. Figure 3 depicts

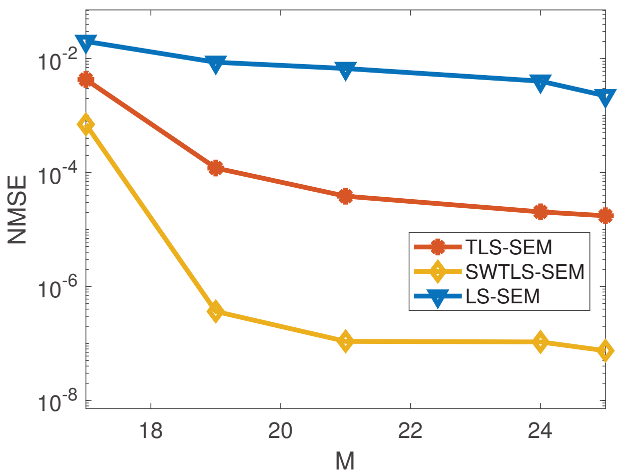

Figure 3: NMSE versus number of samples

the performance of LS-SEM, TLS-SEM and SWTLS-SEM in terms of normalized mean-square error (NMSE), with $\text{NMSE} := \sum_{ij}([\bar{\mathbf{D}}_S\hat{Y}]_{ij} - [\bar{\mathbf{D}}_S Y]_{ij})^2/\sum_{ij}[\bar{\mathbf{D}}_S Y]_{ij}^2$, where $\bar{\mathbf{D}}_S$ denotes the complement of the selection matrix. Figure 3 compares three methods, namely the LS-SEM, TLS-SEM and SWTLS-SEM, as a function of the number of samples M.

As expected intuitively, estimation performance improves considerably as extra prior information is accounted for.

VI. Conclusions and Research Outlook

Two major learning tasks over graphs were considered in this paper in the presence of perturbations. With model mismatch, error-prone laboratory measurements and outages of physical networks, the need arises to account for perturbations in the signal reconstruction and topology inference tasks that in this paper were addressed using approaches based on total least-squares and structural equation models (TLS-SEMs). Structured and weighted (SW) variants of TLS-SEMs leverage prior information to improve performance. Numerical tests on simulated data demonstrated the efficacy of the proposed algorithms.

In addition to thorough experimentation and comparisons with real data sets, future research directions include distributed TLS-SEM approaches to accommodate large-scale networks; more general SEM models to cope with data exhibiting nonlinear dependencies and/or dynamic behaviors; and algorithms to learn how to propagate labels in the presence of perturbations.

References

[1] 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.

[2] 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.

[3] P. Balachandran, E. D. Kolaczyk, and W. D. Viles, "On the propagation of low-rate measurement error to subgraph counts in large networks," *Journal of Machine Learning Research*, vol. 18, no. 1, pp. 2025–2057, July 2017.

[4] 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, Dec. 2013.

[5] 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.

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

[7] ——, "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, July 2018, pp. 1–5.

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

[9] 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.

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

[11] 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, May 2018.

[12] V. N. Ioannidis, Y. Shen, and G. B. Giannakis, "Semi-blind inference of topologies and dynamical processes over graphs," *arXiv preprint arXiv:1805.06095*, May 2018.

[13] 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.

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

[15] 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.

[16] 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.

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

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

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

[20] 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.

[21] 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.

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

[23] 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.

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