

Article

Joint Topology Learning and Latent Input Identification Using Spatio-Temporally Linear Structured SEM

Jie Zhou , Rui Yang, Xintong Shi and Shuyang Feng 

College of Information Science and Engineering, Hohai University, Changzhou 213200, China

* Correspondence: jiezhou@hhu.edu.cn

Abstract

Topology identification and signal inference are cornerstone tasks in graph signal processing (GSP). Structural Equation Modeling (SEM) is particularly effective for network inference as it explicitly captures causal dependencies. However, a major bottleneck in existing SEM-based approaches is the reliance on fully observable exogenous inputs. In many practical applications, systems are driven by latent stimuli, rendering traditional estimation methods ineffective. To overcome this, we propose a novel SEM framework for the joint inference of graph topology and unknown exogenous inputs. The core innovation lies in the spatio-temporal modeling of these latent inputs: each stimulus is decomposed into a rank-one component characterized by nodal sparsity (spatial localization) and temporal piecewise smoothness (temporal persistence). This structured formulation transforms an otherwise ill-posed blind identification problem into a tractable regularized optimization task. We develop an efficient algorithm based on the Alternating Direction Method of Multipliers (ADMM) to solve the resulting convex problem. Numerical experiments on synthetic and real-world datasets demonstrate that the proposed method effectively disentangles endogenous network interactions from latent exogenous influences, outperforming baseline approaches in both topology and signal recovery.

Keywords: Alternating Direction Method of Multipliers (ADMM); blind identification; spatio-temporal modeling; network topology inference; Structural Equation Modeling (SEM); sparse reconstruction

MSC: 62J05; 05C50; 94A12

1. Introduction

Graph signal processing (GSP) [1–5] offers a principled framework for analyzing data on non-Euclidean domains by representing signals on graph nodes and dependencies through edges. This structured formulation enables the study of complex networks—such as social systems—where GSP has proven effective in uncovering patterns in information diffusion, opinion dynamics, and community structures.

Among the core tasks in graph-based learning are topology identification and signal inference. Structural Equation Modeling (SEM) [6–8] has become a foundational tool in this setting, particularly for network analysis. Its strength lies in modeling directed relationships between nodes, thereby distinguishing direct interactions from indirect dependencies—addressing a critical limitation of correlation-based methods. In this work, we focus on linear SEMs, which characterize nodal relationships through a set of linear equations. This linear assumption is justified in many large-scale networked systems where



Academic Editor: Ping Lang

Received: 26 January 2026

Revised: 24 February 2026

Accepted: 26 February 2026

Published: 1 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

signals fluctuate near a steady-state equilibrium or where the coupling between agents is relatively weak. Consequently, linear SEM has been widely applied to network topology identification across various domains including effective connectivity mapping in brain networks [9], monitoring of distributed sensor systems [10], and causal analysis of financial time-series [11–17].

The evolution of SEM estimation techniques reflects a continuous pursuit of computational efficiency and scalability. Early foundational works by Jöreskog [8] and Bentler [7] primarily relied on Maximum Likelihood (ML) estimation. While statistically rigorous, these methods often entailed high computational costs and strict distributional assumptions. To address these issues, Generalized Least Squares (GLS) and Two-Stage Least Squares (2SLS) estimators were subsequently developed by Browne [18] and Bollen [19], offering more computationally tractable alternatives. Building on these LS-based foundations, recent advances by Cai et al. [13] and Baingana et al. [12] integrated sparsity-inducing regularization (e.g., ℓ_1 -norm) into the SEM framework. This crucial innovation enabled the effective identification of high-dimensional topologies, such as gene regulatory networks and social cascades, from limited data.

However, a critical limitation persists even in these state-of-the-art approaches. While robust variations (e.g., TLS-SEM) have been developed to mitigate measurement noise [20,21], they fundamentally rely on the availability of observed input regressors. Consequently, existing frameworks typically operate under the restrictive assumption that exogenous inputs are either fully observable (albeit noisy) or modeled merely as white noise. In many practical scenarios, however, systems are driven by latent or unmeasured stimuli. For instance, in brain connectivity analysis, neural activity is often triggered by sensory inputs not explicitly captured by scanning modalities [22]; similarly, in social networks, information cascades may be initiated by unobserved external events [23]. When applied to such data, existing methods like [12,13] fail to distinguish between internal causal edges and external driving forces, leading to biased or erroneous topology estimates. Effectively addressing this limitation requires solving a blind identification problem—a challenging task recently explored in the context of graph filters [24], but which remains under-addressed in causal SEM frameworks.

To tackle this challenge, we propose a novel SEM-based model that jointly infers network topology and the structure of unknown exogenous inputs within a unified optimization framework. Drawing inspiration from sparse component analysis [25] and dictionary learning for graph signals [26], we model the latent inputs as structured components rather than random noise. Specifically, our model explicitly accounts for the active nodes and temporal duration of external influences by enforcing spatio-temporal sparsity constraints, such as the fused lasso penalty for piecewise temporal smoothness [27]. This structured formulation enables the model to capture complex system dynamics while remaining robust under realistic assumptions. To solve the resulting convex optimization problem efficiently, we develop an algorithm based on the Alternating Direction Method of Multipliers (ADMM) [28–31].

The remainder of this paper is structured as follows. Section 2 reviews the related work in graph signal processing and Structural Equation Modeling. Section 3 introduces the proposed algorithm, including the spatio-temporal SEM model, the MAP-based problem formulation, and a theoretical discussion on identifiability and well-posedness. It also details the ADMM-based numerical method developed to solve the resulting convex optimization task. Section 4 presents comprehensive numerical experiments on both synthetic and real-world datasets to validate the performance of our approach. Finally, Section 5 concludes the paper and discusses potential future research directions.

2. Related Work

Current approaches to graph topology inference can be broadly categorized by their modeling assumptions, ranging from classical estimation with known inputs to robust methods handling measurement noise, and ultimately to blind identification strategies.

2.1. Sparse Structural Equation Modeling

The classical SEM framework, rooted in psychometrics and econometrics, originally relied on ML estimation [7,8]. While asymptotically efficient, ML estimators involve convex optimization and scale poorly with network size, rendering them impractical for high-dimensional signal processing.

To mitigate the curse of dimensionality, recent advances have integrated compressive sensing principles into the SEM framework. By assuming a sparse underlying topology, researchers have formulated inference as a convex optimization problem regularized by the ℓ_1 -norm. For instance, discrete-time approaches [12] and their vector autoregressive (VAR) extensions [13] successfully recover directed edges by minimizing least-squares (LS) prediction errors.

A fundamental weakness of these sparse LS methods is their reliance on the perfect information assumption. They treat the exogenous input matrix $\mathbf{X}$ as ground truth. In practice, inputs are often corrupted or only partially observed. When the regressor $\mathbf{X}$ contains noise, standard LS estimators minimize a biased objective, leading to the identification of spurious edges and asymptotic inconsistency.

2.2. Robust Inference Under Measurement Uncertainty

Acknowledging that real-world data is rarely pristine, subsequent research shifted toward robust topology identification via an “errors-in-variables” perspective. This approach acknowledges that both the dependent variables (network outputs $\mathbf{Y}$) and the independent variables (external inputs $\mathbf{X}$) are subject to perturbations.

The TLS-SEM framework [20,21] was introduced to jointly correct the noisy input regressors and estimate the adjacency matrix. By minimizing the Frobenius norm of the perturbation matrix alongside the fitting error, TLS-SEM provides a statistically consistent estimator under Gaussian noise assumptions. Further extensions, such as structured TLS (sTLS) [20], incorporate sparsity constraints to simultaneously handle input uncertainty and topological sparsity.

While TLS methods mitigate measurement noise, they are correctional rather than generative. They operate on the premise that a baseline observation of the input exists. Consequently, these methods falter in “blind” scenarios where driving stimuli are latent (fully unobserved). Treating a missing input as “zero plus Gaussian noise” is mathematically ill-posed and fails to capture the inherent spatio-temporal structure (e.g., block sparsity) of real-world latent drivers.

2.3. Blind Identification and Joint Recovery

The challenge of identifying system dynamics driven by unobserved inputs has been extensively explored in the domain of graph filter identification. Approaches in this category [24,26] leverage the assumption that graph signals are smooth or sparse with respect to underlying topology, typically solving a joint matrix factorization problem to recover the Graph Shift Operator (GSO) and sparse inputs simultaneously.

However, graph filter-based methods frequently assume that the GSO is symmetric (undirected) or that the filter coefficients commute with the adjacency matrix. These spectral assumptions are too restrictive for causal SEMs, which must model directed, asymmetric information flows (e.g., $i \rightarrow j$ does not imply $j \rightarrow i$).

The existing literature presents a clear dichotomy: SEM-based methods [12,20] offer causal interpretability but demand observed inputs, while blind graph filter methods [24] handle latent inputs but often sacrifice the directionality inherent to SEMs. Our work bridges this gap by proposing a blind SEM framework that enforces spatio-temporal sparsity on latent inputs, thereby preserving causal directionality without requiring prior input measurements.

3. Proposed Algorithm

3.1. SEM Model

Consider a directed network of N nodes represented by an adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$, where an entry $A_{i,j} \neq 0$ denotes a directed edge from node j to node i . We assume the graph contains no self-loops, such that $A_{i,i} = 0$ for all i . Let $y_i(t)$ denote the observation at node i at time t , which is modeled as a linear combination of measurements from its incoming neighbors and an exogenous input $x_i(t)$

$$y_i(t) = \sum_{j \neq i} A_{i,j} y_j(t) + x_i(t), \quad t = 1, \dots, T. \quad (1)$$

By defining the observation vector $\mathbf{y}(t) = [y_1(t), \dots, y_N(t)]^\top$ and the input vector $\mathbf{x}(t) = [x_1(t), \dots, x_N(t)]^\top$, the system can be expressed in compact form as $\mathbf{y}(t) = \mathbf{A}\mathbf{y}(t) + \mathbf{x}(t)$. Over T time steps, the relationship is captured by the SEM

$$\mathbf{Y} = \mathbf{A}\mathbf{Y} + \mathbf{X}, \quad (2)$$

where $\mathbf{Y} = [\mathbf{y}(1), \dots, \mathbf{y}(T)] \in \mathbb{R}^{N \times T}$ and $\mathbf{X} = [\mathbf{x}(1), \dots, \mathbf{x}(T)] \in \mathbb{R}^{N \times T}$ are the observation and exogenous input matrices, respectively.

It should be noted that the relationship in (2) describes the system's response to latent stimuli across the entire observation window. While a steady-state analysis could simplify the estimation of $\mathbf{A}$, our framework is specifically designed to capture the transient dynamics and temporal evolution of $\mathbf{X}$. Regarding the initial conditions, we assume that the system is either observed from a state of rest or that the effect of the initial state $\mathbf{y}(0)$ is encapsulated within the identified latent component $\mathbf{X}$ during the initial time steps, ensuring the tractability of the blind identification problem.

The structural equation in (2) implies a mapping $\mathbf{Y} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{X}$. For the inferred network to be physically plausible and stable, the matrix $(\mathbf{I} - \mathbf{A})$ must be invertible, which is generally satisfied when the spectral radius of the adjacency matrix follows $\rho(\mathbf{A}) < 1$. In the context of real-world networks (e.g., neural or sensor networks), this condition is naturally maintained by the inherent energy dissipation and negative feedback mechanisms within the system. While our optimization framework in (5a) does not explicitly enforce a spectral constraint to preserve convexity, the row-sparsity and no-self-loop ($A_{i,i} = 0$) priors effectively regularize the eigenvalues of $\mathbf{A}$, ensuring that the identified topology represents a stable dynamical system.

Our formulation is guided by two primary assumptions regarding the latent exogenous inputs:

- A single exogenous stimulus may affect a subset of nodes, and a single node may be influenced by multiple distinct stimuli.
- Each stimulus typically persists over a specific duration, with onset and offset times that vary across different inputs.

Based on these assumptions, the exogenous input matrix is modeled as a sum of Q rank-one components plus stochastic noise:

$$\mathbf{X} = \sum_{q=1}^Q \mathbf{d}_q \mathbf{c}_q^\top + \mathbf{E}, \quad (3)$$

where $Q \ll N$ represents the total number of latent inputs and $\mathbf{E} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ denotes the i.i.d. Gaussian noise. The vector $\mathbf{d}_q \in \mathbb{R}^N$ identifies the spatial profile (excited nodes), while $\mathbf{c}_q \in \mathbb{R}^T$ represents the temporal evolution of the q -th input. We emphasize that while $\mathbf{E}$ represents unpredictable noise, the structured component $\sum \mathbf{d}_q \mathbf{c}_q^\top$ captures the deterministic driving forces external to the network's internal dynamics.

Because exogenous inputs generally persist over continuous intervals, $\mathbf{c}_q$ is expected to exhibit piecewise constant behavior. This is characterized by the first-order temporal difference operator $\mathbf{D}_T \in \mathbb{R}^{T \times (T-1)}$, defined as the bidiagonal matrix

$$\mathbf{D}_T = \begin{bmatrix} -1 & & & \\ 1 & -1 & & \\ & 1 & \ddots & \\ & & \ddots & -1 \\ & & & 1 \end{bmatrix}, \quad (4)$$

where the sparsity of the vector $\mathbf{c}_q^\top \mathbf{D}_T$ reflects the temporal persistence of the q -th input.

Given that $\mathbf{X}$ is the sum of Q rank-one matrices, its rank is at most Q . To ensure that the identification of $\mathbf{X}$ is tractable, we employ nuclear norm regularization $\|\mathbf{X}\|_*$ to promote a low-rank structure. Furthermore, since each spatial profile $\mathbf{d}_q$ is sparse, $\mathbf{X}$ inherently exhibits row-sparsity, which motivates the use of an $\ell_{2,1}$ -norm penalty to encourage joint sparsity across the rows of the input matrix.

It is important to clarify that while the SEM in (2) assumes instantaneous dependencies, it is closely related to dynamic models such as vector autoregressive (VAR) processes. The instantaneous formulation captures the contemporaneous causal effects occurring within a single sampling interval, which is particularly relevant in systems where information propagation is faster than the measurement frequency. By enforcing temporal consistency on $\mathbf{X}$, we bridge the gap between a static network structure and the dynamic nature of graph signals.

3.2. Problem Formulation

In practice, observations in the SEM may be corrupted by additive noise. To provide a rigorous statistical foundation, we formulate the joint identification task within a maximum a posteriori (MAP) estimation framework. Assuming the observation noise follows a Gaussian distribution and the structural priors (sparsity and smoothness) follow corresponding Laplacian or fused-lasso distributions, the MAP estimator is equivalent to solving the following regularized optimization problem:

$$\min_{\mathbf{A}, \mathbf{X}} \|\mathbf{Y} - \mathbf{A}\mathbf{Y} - \mathbf{X}\|_F^2 + \alpha \|\mathbf{A}\|_1 + \beta \|\mathbf{X}\|_* + \gamma g_\delta(\mathbf{X}) \quad (5a)$$

$$A_{i,i} = 0, \quad i = 1, \dots, N, \quad (5b)$$

where the constraint $A_{i,i} = 0$ excludes self-loops in the adjacency matrix $\mathbf{A}$. Here, α , β , and γ are regularization coefficients. The penalty function $g_\delta(\mathbf{X})$ is defined as a convex combination of two structural priors

$$g_\delta(\mathbf{X}) = (1 - \delta)\|\mathbf{X}\|_{2,1} + \delta\|\mathbf{X}\mathbf{D}_T\|_1 \quad (6)$$

with $0 \leq \delta \leq 1$. In this formulation, $\|\mathbf{X}\|_{2,1} = \sum_{i=1}^N \|\mathbf{x}_i\|_2$ represents the $\ell_{2,1}$ -norm, where $\mathbf{x}_i^\top$ denotes the i -th row of $\mathbf{X}$. The first term in $g_\delta(\mathbf{X})$ is introduced to promote row-sparsity, while the second term is used to promote temporal consistency.

The inclusion of these specific regularization terms is justified by the following structural priors:

- $\|\mathbf{A}\|_1$ promotes a sparse network topology, consistent with the belief that most real-world networks are characterized by a relatively small number of direct causal links to highlight the major dependences.
- $\|\mathbf{X}\|_*$ encourages a low-rank structure in $\mathbf{X}$, reflecting the assumption that the network dynamics are driven by a limited number of latent stimuli.
- $\|\mathbf{X}\|_{2,1}$ facilitates row-sparsity to identify specific nodes that serve as the primary entry points for exogenous influences.
- $\|\mathbf{X}\mathbf{D}_T\|_1$ enforces piecewise temporal smoothness, ensuring that the identified inputs persist over meaningful durations rather than manifesting as transient noise.

It is important to distinguish the estimation principle from numerical optimization. While the MAP framework motivates the objective function in (5a), the Alternating Direction Method of Multipliers (ADMM) is employed as the computational tool to efficiently solve this convex optimization problem. By decomposing the joint identification task into smaller, tractable sub-problems, the ADMM-based solver ensures numerical stability and provides reliable estimates of the network topology $\mathbf{A}$ and latent stimuli $\mathbf{X}$ that are consistent with the prescribed structural priors.

While expectation maximization (EM) is a common strategy for models with latent variables, we adopt a direct optimization approach via ADMM. This choice is motivated by the presence of multiple non-smooth regularizers (e.g., ℓ_1 and nuclear norms), for which ADMM provides efficient proximal updates. Furthermore, given the convexity of the formulated MAP problem, the ADMM-based solver ensures global convergence and higher computational efficiency in high-dimensional settings compared to the iterative E-steps required in EM.

Discussion on Identifiability and Well-Posedness

The joint recovery of the adjacency matrix $\mathbf{A}$ and the latent input matrix $\mathbf{X}$ from observations $\mathbf{Y} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{X}$ is inherently a blind identification problem, which immediately raises the question of identifiability: can different pairs $(\mathbf{A}, \mathbf{X})$ produce exactly the same observations $\mathbf{Y}$? Such observational equivalence would render the decomposition non-unique and the estimation ill-posed [32]. In our framework, however, the potential for ambiguity is significantly mitigated by the non-overlapping structural constraints imposed on the two variables.

To understand this, consider any hypothetical linear transformation $\mathbf{M}$ that attempts to reparameterize the system as

$$\mathbf{A}' = \mathbf{A} + \mathbf{M}, \quad \mathbf{X}' = \mathbf{X} - \mathbf{M}\mathbf{Y},$$

while preserving the observation equation $\mathbf{Y} = \mathbf{A}'\mathbf{Y} + \mathbf{X}'$. For such a transformation to be observationally equivalent, it must satisfy $\mathbf{A}'\mathbf{Y} + \mathbf{X}' = \mathbf{A}\mathbf{Y} + \mathbf{X}$. The key observation is that any non-trivial $\mathbf{M}$ inevitably violates at least one of the prescribed structural priors:

- **Violation of topological constraints:** The adjacency matrix $\mathbf{A}$ is constrained to have zero diagonal entries ($A_{i,i} = 0$) and is regularized by the ℓ_1 norm to encourage sparsity. If a transformation attempts to absorb part of $\mathbf{X}$ into $\mathbf{A}$ (i.e., $\mathbf{M} \neq \mathbf{0}$), the resulting $\mathbf{A}'$ would generally acquire non-zero diagonal elements or become densely populated, directly conflicting with the no-self-loop constraint and incurring a large ℓ_1 penalty.
- **Violation of input structure:** Conversely, transferring information from $\mathbf{A}$ into $\mathbf{X}$ would disrupt the low-rank and piecewise smooth temporal structure of $\mathbf{X}$. The nuclear norm penalty $\|\mathbf{X}\|_*$ would increase because the rank of $\mathbf{X}'$ would likely exceed the true number of latent sources Q . Moreover, the temporal consistency term $\|\mathbf{X}\mathbf{D}_T\|_1$ would penalize any newly introduced abrupt changes that are not characteristic of the actual exogenous stimuli.

Thus, the joint spatio-temporal and low-rank regularizers act as a mathematical filter that enforces a unique decomposition by ensuring that $\mathbf{A}$ and $\mathbf{X}$ reside in disjoint low-complexity manifolds [25]. Incoherence between these manifolds—one promoting sparse, diagonal-free matrices, the other promoting low-rank and piecewise constant row structures—guarantees that any mixing of the two components would be heavily penalized by the objective function in (5a).

Beyond this qualitative argument, the well-posedness of the problem can also be examined through a degrees-of-freedom perspective. The number of unknown parameters in $\mathbf{A}$ (non-zero edges s) and in the structured part of $\mathbf{X}$ (the low-rank factorization $Q(N + T)$) must be small relative to the total number of observations $N \times T$. Concretely, for the inverse problem to be well-conditioned, the total effective degrees of freedom $s + Q(N + T)$ should be less than NT . The regularization parameters α, β, γ in (5a) serve as “complexity controllers” that automatically balance this trade-off during optimization.

A rigorous theoretical proof of identifiability under general conditions remains an open challenge, as is common in blind source separation, dictionary learning, and robust principal component analysis (RPCA) [25]. Nevertheless, the structural disparity between the two components, combined with the convex regularization framework, provides strong empirical guarantees. The stable recovery results reported in Sections 4.1.2 and 4.1.3 (e.g., the monotonic improvement with increasing sample size T and the high F1-scores in topology recovery) confirm that the proposed decomposition is identifiable under realistic spatio-temporal assumptions.

3.3. Numerical Method

While expectation maximization (EM) is a conventional choice for latent variable estimation, it can be computationally intensive when the model incorporates multiple non-smooth regularizers. In this work, we adopt the Alternating Direction Method of Multipliers (ADMM), as it allows for the seamless integration of proximal mappings for the ℓ_1 , nuclear, and $\ell_{2,1}$ norms, providing a more direct and efficient computational path for high-dimensional topology inference.

The problem formulated above is convex and can be tackled by well-established solvers. To improve computational efficiency and scalability for large-scale networks, we develop a customized framework based on the Alternating Direction Method of Multipliers (ADMM). We first introduce an auxiliary variable $\mathbf{Z} = \mathbf{X}$ to decouple the low-rank and row-sparse regularizers. The augmented Lagrangian is given by

$$\begin{aligned} \mathcal{L}_\rho(\mathbf{A}, \mathbf{X}, \mathbf{Z}, \Theta) &= \|\mathbf{Y} - \mathbf{A}\mathbf{Y} - \mathbf{X}\|_F^2 + \alpha\|\mathbf{A}\|_1 + \beta\|\mathbf{Z}\|_* + \gamma g_\delta(\mathbf{X}) \\ &+ \text{Tr}\left\{\Theta^\top(\mathbf{X} - \mathbf{Z})\right\} + \frac{\rho}{2}\|\mathbf{X} - \mathbf{Z}\|_F^2, \end{aligned} \quad (7)$$

where Θ is the matrix of Lagrange multipliers corresponding to the constraint $\mathbf{X} = \mathbf{Z}$ and $\rho > 0$. The ADMM alternately updates $\mathbf{A}$, $\mathbf{X}$, $\mathbf{Z}$, and Θ , while fixing the other variables. The major steps are summarized below.

3.3.1. Update of $\mathbf{A}$

The adjacency matrix $\mathbf{A}$ is updated by

$$\mathbf{A}^{(k+1)} = \arg \min_{\mathbf{A}} \left\| \mathbf{A}\mathbf{Y} - \mathbf{R}^{(k)} \right\|_F^2 + \alpha \|\mathbf{A}\|_1, \quad \text{s.t. } A_{i,i} = 0 \quad (8)$$

where $\mathbf{R}^{(k)} = \mathbf{Y} - \mathbf{X}^{(k)}$. Each row $\mathbf{a}_i^\top$ of $\mathbf{A}$ (excluding the diagonal element $A_{i,i}$) can be updated in parallel:

$$\mathbf{a}_i^{(k+1)} = \arg \min_{\mathbf{a} \in \mathbb{R}^{N-1}} \left\| \mathbf{r}_i^{(k)} - \mathbf{Y}_{-i}^\top \mathbf{a} \right\|_2^2 + \alpha \|\mathbf{a}\|_1, \quad (9)$$

where $\mathbf{Y}_{-i}$ is $\mathbf{Y}$ with its i -th row removed and $\left(\mathbf{r}_i^{(k)}\right)^\top$ denotes the i -th row of $\mathbf{R}^{(k)}$. This is the standard Lasso model [33]. We employ the fast iterative shrinkage thresholding algorithm (FISTA) to solve it [34]. The main steps are summarized as follows:

$$\mathbf{b}_i^{(t)} = \mathbf{a}_i^{(t-1)} + \frac{\eta^{(t-1)} - 1}{\eta^{(t)}} \left(\mathbf{a}_i^{(t-1)} - \mathbf{a}_i^{(t-2)} \right), \quad (10a)$$

$$\mathbf{a}_i^{(t)} = \mathcal{S}_{\alpha/L} \left(\mathbf{b}_i^{(t)} - \frac{2}{L} \mathbf{Y}_{-i} \left(\mathbf{Y}_{-i}^\top \mathbf{b}_i^{(t)} - \mathbf{r}_i^{(k)} \right) \right), \quad (10b)$$

$$\eta^{(t+1)} = \frac{1 + \sqrt{1 + 4(\eta^{(t)})^2}}{2}. \quad (10c)$$

In these updates, t denotes the inner iteration index with initial values $\mathbf{a}_i^{(-1)} = \mathbf{a}_i^{(0)} = \mathbf{a}_i^{(k)}$ and $\eta^{(0)} = 1$. The constant $L \geq 2\lambda_{\max}(\mathbf{Y}_{-i} \mathbf{Y}_{-i}^\top)$ represents the Lipschitz constant of the gradient of the smooth least-squares loss, where $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue operator. The soft-thresholding operator $\mathcal{S}_\tau(x) = \text{sgn}(x) \cdot \max(|x| - \tau, 0)$ is the proximal operator of the ℓ_1 -norm, which induces the topological sparsity in $\mathbf{A}$.

3.3.2. Update of $\mathbf{X}$

The input matrix $\mathbf{X}$ is updated by

$$\mathbf{X}^{(k+1)} = \arg \min_{\mathbf{X}} \left\{ \left\| \mathbf{B}^{(k)} - \mathbf{X} \right\|_F^2 + \gamma \mathcal{G}_\delta(\mathbf{X}) + \frac{\rho}{2} \left\| \mathbf{X} - \Delta^{(k)} \right\|_F^2 \right\} \quad (11)$$

where $\mathbf{B}^{(k)} = \mathbf{Y} - \mathbf{A}^{(k+1)}\mathbf{Y}$ and $\Delta^{(k)} = \mathbf{Z}^{(k)} - \frac{1}{\rho}\Theta^{(k)}$. By decomposing this problem row-wise, for each row $\mathbf{x}_i^\top$ we solve the following optimization problem

$$\mathbf{x}_i^{(k+1)} = \arg \min_{\mathbf{x} \in \mathbb{R}^T} \left\| \mathbf{x} - \mathbf{c}_i^{(k)} \right\|_2^2 + \gamma' \|\mathbf{x}\|_2 + \delta' \left\| \mathbf{D}_T^\top \mathbf{x} \right\|_1 \quad (12)$$

where $\mathbf{c}_i^{(k)} = \frac{2\mathbf{b}_i^{(k)} + \rho\mathbf{d}_i^{(k)}}{2+\rho}$, and the scaled regularization parameters are $\gamma' = \frac{2\gamma(1-\delta)}{2+\rho}$ and $\delta' = \frac{2\gamma\delta}{2+\rho}$. Here, $\left(\mathbf{b}_i^{(k)}\right)^\top$ and $\left(\mathbf{d}_i^{(k)}\right)^\top$ denote the i -th row of $\mathbf{B}^{(k)}$ and $\Delta^{(k)}$, respectively.

We employ the Primal-Dual Hybrid Gradient (PDHG) algorithm [35] to solve this subproblem. By introducing the dual variable $\mathbf{g} \in \mathbb{R}^{T-1}$, the ℓ_1 norm term is reformulated

via its convex conjugate as $\max_{\|\mathbf{g}\|_\infty \leq \delta'} (\mathbf{g}^\top \mathbf{D}_T^\top \mathbf{x})$. The subproblem is thus equivalent to finding the saddle point of the following Lagrangian:

$$\min_{\mathbf{x}} \max_{\|\mathbf{g}\|_\infty \leq \delta'} \left\| \mathbf{x} - \mathbf{c}_i^{(k)} \right\|_2^2 + \gamma' \|\mathbf{x}\|_2 + \mathbf{g}^\top \mathbf{D}_T^\top \mathbf{x}. \quad (13)$$

The saddle-point problem is solved iteratively using the Primal-Dual Hybrid Gradient (PDHG) algorithm with an inner iteration index m . At each iteration, the variables are updated using the dual and primal step sizes, denoted by $\sigma > 0$ and $\tau > 0$, respectively. To guarantee the convergence of the algorithm, these parameters are chosen to satisfy the stability condition $\sigma\tau L^2 < 1$, where $L = \|\mathbf{D}_T^\top\|_2$ denotes the operator norm of the difference matrix $\mathbf{D}_T^\top$. Furthermore, the dual update involves an orthogonal projection onto the ℓ_∞ -ball with radius δ' , which is denoted by $\text{proj}_{\|\cdot\|_\infty \leq \delta'}(\cdot)$.

- Dual Update (Projection Step):

$$\mathbf{g}^{(m+1)} = \text{proj}_{\|\cdot\|_\infty \leq \delta'} \left(\mathbf{g}^{(m)} + \sigma \mathbf{D}_T^\top \bar{\mathbf{x}}^{(m)} \right). \quad (14)$$

This is performed via element-wise clamping: $g_j = \text{sgn}(q_j) \cdot \min(|q_j|, \delta')$, where $\mathbf{q} = \mathbf{g}^{(m)} + \sigma \mathbf{D}_T^\top \bar{\mathbf{x}}^{(m)}$. Here, $\bar{\mathbf{x}}^{(m)}$ is the extrapolated primal variable used to ensure convergence stability.

- Primal Update (Proximal Step): The update for $\mathbf{x}$ combines the quadratic fidelity term and the ℓ_2 penalty. First, an intermediate vector $\hat{\mathbf{v}}$ is computed by solving the optimality condition of the quadratic part of the Lagrangian:

$$\hat{\mathbf{v}} = \frac{2\tau \mathbf{c}_i^{(k)} + \mathbf{x}^{(m)} - \tau \mathbf{D}_T \mathbf{g}^{(m+1)}}{1 + 2\tau}. \quad (15)$$

Subsequently, the primal variable $\mathbf{x}^{(m+1)}$ is obtained by applying the ℓ_2 proximal operator (block soft-thresholding) to $\hat{\mathbf{v}}$:

$$\mathbf{x}^{(m+1)} = \frac{\hat{\mathbf{v}}}{\|\hat{\mathbf{v}}\|_2} \max \left(\|\hat{\mathbf{v}}\|_2 - \frac{\tau\gamma'}{1 + 2\tau}, 0 \right). \quad (16)$$

- Extrapolation Step: To stabilize the interaction between primal and dual variables, the extrapolated variable $\bar{\mathbf{x}}$ is updated for the next dual step:

$$\bar{\mathbf{x}}^{(m+1)} = \mathbf{x}^{(m+1)} + \theta \left(\mathbf{x}^{(m+1)} - \mathbf{x}^{(m)} \right) \quad (17)$$

where $\theta \in [0, 1]$ is the relaxation parameter used in the extrapolation step to stabilize the interaction between primal and dual variables, ensuring the algorithm converges to the saddle point of the subproblem.

3.3.3. Update of $\mathbf{Z}$

The auxiliary variable $\mathbf{Z}$ is updated by

$$\begin{aligned} \mathbf{Z}^{(k+1)} &= \arg \min_{\mathbf{Z}} \frac{1}{2} \left\| \mathbf{Z} - \tilde{\mathbf{X}}^{(k)} \right\|_F^2 + \frac{\beta}{\rho} \|\mathbf{Z}\|_* \\ &= \mathcal{D}_{\frac{\beta}{\rho}} \left(\tilde{\mathbf{X}}^{(k)} \right) \end{aligned} \quad (18)$$

where $\tilde{\mathbf{X}}^{(k)} = \mathbf{X}^{(k+1)} + \frac{1}{\rho} \Theta^{(k)}$, and $\mathcal{D}_\lambda(\mathbf{X})$ is the singular-value thresholding operator, given by

$$\mathcal{D}_\lambda(\mathbf{X}) = \mathbf{U} \mathcal{D}_\lambda(\Sigma) \mathbf{V}^\top, \quad (19)$$

with $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ as the singular value decomposition of $\mathbf{X}$, $\mathbf{\Sigma} = \text{diag}(\{\sigma_i\}_{i=1}^r)$ containing singular values σ_i , and $\mathcal{D}_\lambda(\mathbf{\Sigma}) = \text{diag}((\sigma_i - \lambda)_+)$ with $t_+ = \max(0, t)$. The operator $(\cdot)_+$ is defined as the positive part of a real number, such that $(t)_+ = \max(0, t)$. Note that the auxiliary variable $\mathbf{Z}$ is introduced as a proximal surrogate to decouple the low-rank and row-sparse constraints within the ADMM framework.

3.3.4. Update of Θ

The Lagrangian multipliers Θ are updated by

$$\Theta^{(k+1)} = \Theta^{(k)} + \rho(\mathbf{X}^{(k+1)} - \mathbf{Z}^{(k+1)}), \quad (20)$$

where $\rho > 0$ is the penalty parameter of the augmented Lagrangian. It controls the step size of the dual update and determines the weight of the quadratic penalty term added to ensure strong convexity during the variable updates.

The iterative procedure continues until the relative updates of $\mathbf{A}$ and $\mathbf{X}$ between consecutive iterations fall below a predefined threshold. Given that the original optimization problem is convex, the convergence of this ADMM-based framework to a global optimum is theoretically guaranteed. The complete execution of the proposed method is summarized in Algorithm 1, and the algorithm is henceforth referred to as XLS-SEM.

Algorithm 1: Proposed Spatio-temporally Structured SEM (STS-SEM) Algorithm

Input : Observation matrix $\mathbf{Y}$, regularization parameters $\alpha, \beta, \gamma, \delta$, and penalty parameter $\rho > 0$

Output: Estimated adjacency matrix $\mathbf{A}$ and latent input matrix $\mathbf{X}$

```

1 begin
2   Initialize  $\mathbf{A}^{(0)}, \mathbf{X}^{(0)}, \mathbf{Z}^{(0)}$ , and  $\Theta^{(0)}$ 
3   Set iteration counter  $k \leftarrow 0$ 
4   while stopping criteria not met and  $k < k_{\max}$  do
5     /* Step 1: Update Adjacency Matrix  $\mathbf{A}$  (Parallel Lasso) */
6     Compute residual  $\mathbf{R}^{(k)} = \mathbf{Y} - \mathbf{X}^{(k)}$ 
7     for  $i = 1, \dots, N$  (in parallel) do
8       Update row  $\mathbf{a}_i^{(k+1)}$  by solving the Lasso subproblem via FISTA (using
9         Equations (10a)–(10c))
10    /* Step 2: Update Input Matrix  $\mathbf{X}$  (Parallel PDHG) */
11    Compute auxiliary variables  $\mathbf{B}^{(k)} = \mathbf{Y} - \mathbf{A}^{(k+1)}\mathbf{Y}$  and  $\Delta^{(k)} = \mathbf{Z}^{(k)} - \frac{1}{\rho}\Theta^{(k)}$ 
12    for  $i = 1, \dots, N$  (in parallel) do
13      Update row  $\mathbf{x}_i^{(k+1)}$  by solving the convex subproblem via PDHG (using
14        Equations (14)–(17))
15    /* Step 3: Update Auxiliary Variable  $\mathbf{Z}$  (SVT) */
16    Compute  $\tilde{\mathbf{X}}^{(k)} = \mathbf{X}^{(k+1)} + \frac{1}{\rho}\Theta^{(k)}$ 
17    Update  $\mathbf{Z}^{(k+1)} = \mathcal{D}_{\beta/\rho}(\tilde{\mathbf{X}}^{(k)})$  via Singular Value Thresholding
18    /* Step 4: Update Multipliers  $\Theta$  */
19    Update  $\Theta^{(k+1)} = \Theta^{(k)} + \rho(\mathbf{X}^{(k+1)} - \mathbf{Z}^{(k+1)})$ 
20     $k \leftarrow k + 1$ 
21 return  $\mathbf{A}^{(k)}, \mathbf{X}^{(k)}$ 

```

4. Numerical Experiments

In this section, we conduct numerical experiments on both synthetic and real-world data to evaluate the performance of the proposed algorithm. To demonstrate its efficacy in handling latent exogenous stimuli, we benchmark our framework against several state-of-the-art baselines that conventionally require direct input measurements:

- LS-SEM [12,13]: The conventional sparse Structural Equation Model based on least-squares estimation, which utilizes ℓ_1 regularization to recover network topology.
- TLS-SEM and sTLS-SEM [20]: Robust variants designed to account for errors in variables. While TLS-SEM addresses measurement noise, sTLS-SEM further incorporates sparsity regularization on the input matrix $\mathbf{E}$.

Crucially, a defining advantage of the XLS-SEM framework is its blind identification capability. To simulate a realistic scenario where exogenous inputs exist but cannot be perfectly captured, the baseline methods (LS-SEM, TLS-SEM, and sTLS-SEM) are provided with noisy observations of the inputs. In contrast, the proposed XLS-SEM receives no prior information regarding exogenous inputs and must infer the network topology solely from observed output signals. All model regularization parameters are determined via exhaustive grid search.

4.1. Synthetic Data

In this suite of experiments, we evaluate the proposed framework using a synthetic network of $N = 20$ nodes generated via an Erdős-Rényi (ER) random graph with an edge connection probability of 0.2. The system dynamics are driven by Q latent exogenous input sources. To faithfully reflect the spatio-temporal characteristics of real-world stimuli, the inputs are synthesized as follows: for each source q , a spatial signature $\mathbf{d}_q$ is created by randomly selecting P active nodes. The corresponding temporal profile $\mathbf{c}_q$ is modeled as a piecewise constant signal containing up to three non-overlapping active segments with randomized onset times and durations (constrained by $\lfloor T/3 \rfloor$). This construction naturally ensures the sparsity of the temporal finite difference $\mathbf{D}_T \mathbf{c}_q$. Non-zero amplitudes of these latent inputs are independently sampled from a uniform distribution over $[-1, 1]$.

Given the topology $\mathbf{A}$ and the exogenous matrix $\mathbf{X}$, the observed signals are generated as $\mathbf{Z} = \mathbf{Y} + \mathbf{E}$, where $\mathbf{Y} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{X}$ represents the latent clean graph signals and $\mathbf{E}$ denotes i.i.d. Gaussian measurement noise is sampled from $\mathcal{N}(0, \sigma_E^2)$. The noise variance σ_E^2 is calibrated to a prescribed output signal-to-noise ratio (SNR_Y), defined as $\text{SNR}_Y = 10 \log_{10} \left(\frac{\|\bar{\mathbf{y}}\|_2^2}{N \sigma_E^2} \right)$, where $\bar{\mathbf{y}} = \frac{1}{T} \mathbf{Y} \mathbf{1}$ is the temporal mean of the signals.

To assess the robustness of blind identification under measurement uncertainty, we simulate a scenario where baseline methods (LS-SEM, TLS-SEM, and sTLS-SEM) are provided with corrupted input measurements $\tilde{\mathbf{X}} = \mathbf{X} + \mathbf{R}$, where $\mathbf{R}$ is adjusted to achieve a target input SNR_X . Crucially, the proposed XLS-SEM algorithm performs inference in a completely blind manner, ignoring $\tilde{\mathbf{X}}$ and relying solely on the observed output $\mathbf{Z}$.

The performance is averaged over 100 independent Monte Carlo trials for each configuration of SNR_Y , SNR_X , Q , P , and T . We employ two primary metrics for performance evaluation and comparison:

- Mean squared error (MSE) of the adjacency matrix: $\text{MSE}_A = \frac{1}{N^2} \|\mathbf{A} - \hat{\mathbf{A}}\|_F^2$ where $\hat{\mathbf{A}}$ denotes the estimated adjacency matrix, evaluating topology recovery.
- MSE of the signal reconstruction (denoising): $\text{MSE}_Y = \frac{1}{NT} \|\mathbf{Y} - \hat{\mathbf{Y}}\|_F^2$ where $\hat{\mathbf{Y}}$ denotes the estimated output matrix.

4.1.1. Visual Illustration and Convergence

Before analyzing the statistical performance metrics, we first provide a qualitative assessment of the proposed XLS-SEM using a representative trial. The synthetic data was generated based on an ER graph with $N = 20$ nodes and connection probability $p = 0.2$. The observation period was set to $T = 100$. To simulate the spatio-temporal sparsity of the latent drivers, we generated $\mathbf{X}$ with $K = 3$ temporal segments and $P = 3$ active nodes per segment. Both the input and output signal-to-noise ratios were set to $\text{SNR}_Y = \text{SNR}_X = 15$ dB.

Figure 1 visually compares the ground truth and estimated matrices. As observed in Figure 1a, the latent input $\mathbf{X}$ exhibits a distinct block-sparse structure, reflecting the spatio-temporal activation patterns of the exogenous sources. The recovered $\hat{\mathbf{X}}$ (Figure 1b) successfully captures these sparsity patterns and active segments with high fidelity. For visual clarity, a small hard threshold (<0.1) was applied to the estimated sparse matrix to suppress negligible numerical residuals inherent to the iterative solver. Consequently, the recovery error map (Figure 1c) shows minimal discrepancies, validating the effectiveness of the joint ℓ_1 and nuclear norm regularization. Similarly, the second row (Figure 1d–f) demonstrates the signal reconstruction capability; despite the observed signals being corrupted by noise, the estimated $\hat{\mathbf{Y}}$ in Figure 1e closely matches the clean ground truth $\mathbf{Y}$ (Figure 1d), corroborating the denoising properties of our framework.

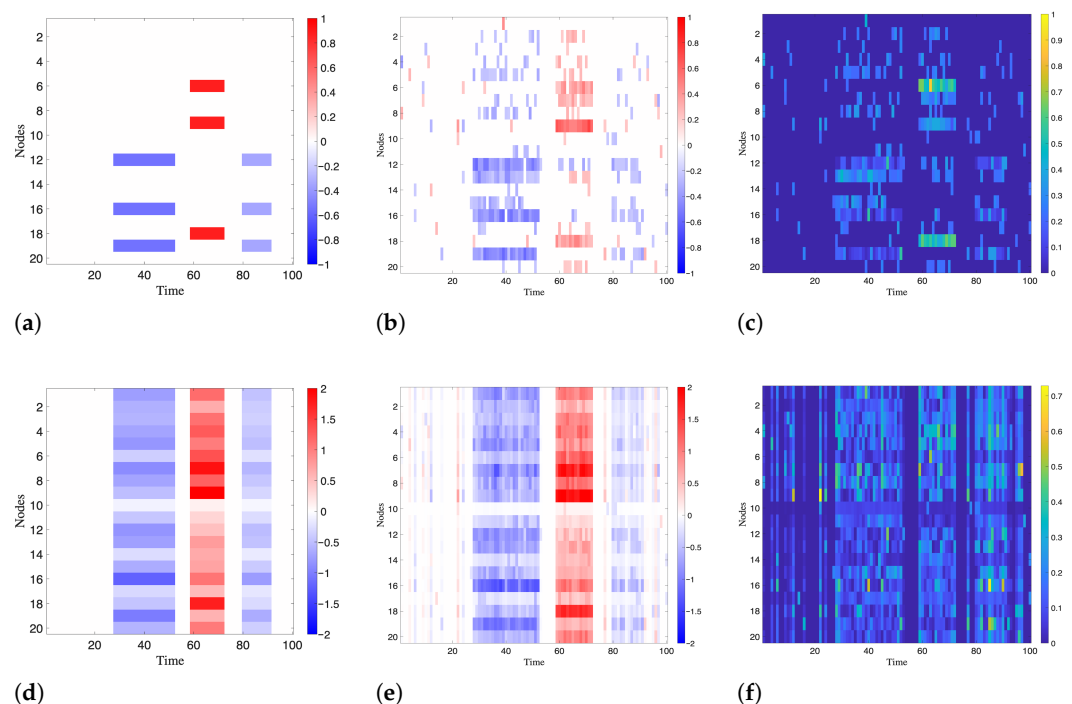


Figure 1. Visual illustration of topology inference and signal recovery on synthetic data ($N = 20$, $T = 100$). (a–c) Input recovery performance: (a) ground truth latent input $\mathbf{X}$; (b) estimated input $\hat{\mathbf{X}}$ (proposed); (c) recovery error $|\mathbf{X} - \hat{\mathbf{X}}|$. (d–f) Signal reconstruction performance: (d) clean graph signal $\mathbf{Y}$; (e) reconstructed signal $\hat{\mathbf{Y}}$ (proposed); (f) reconstruction error $|\mathbf{Y} - \hat{\mathbf{Y}}|$.

The convergence behavior of the ADMM-based solver is illustrated in Figure 2. The plot displays the relative change of the updated variables $\mathbf{A}$ and $\mathbf{X}$ versus the iteration count on a logarithmic scale. Both curves exhibit a rapid and monotonic decay, dropping below the tolerance threshold (10^{-4}) within a few hundred iterations. This empirical evidence confirms the computational efficiency and numerical stability of the proposed algorithm.

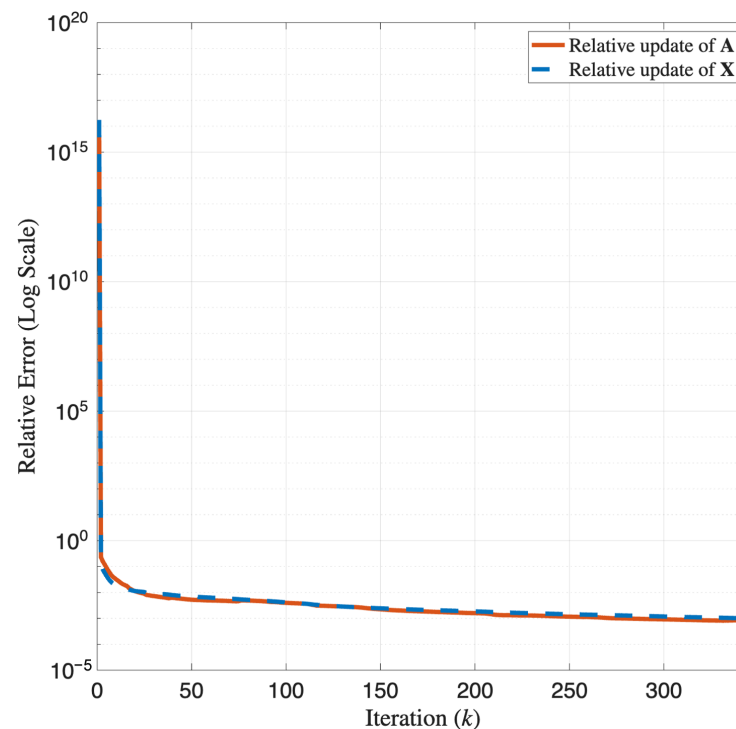


Figure 2. Convergence analysis of the proposed ADMM solver. The curves show the relative change of variables $\mathbf{A}$ and $\mathbf{X}$ over iterations.

4.1.2. Impact of Input and Output SNR

We first examine the sensitivity of the algorithms to noise levels by varying SNR_Y and SNR_X independently, while fixing the non-varying parameter at 15 dB. The remaining parameters are chosen as $Q = 2$, $P = 3$, and $T = 100$. Figure 3a,b illustrate topology recovery and reconstruction accuracy under varying observation noise (SNR_Y), while Figure 3c,d present the results for varying input noise (SNR_X).

As anticipated, the performance of all methods improves monotonically with increasing SNR. Notably, XLS-SEM consistently outperforms the baseline methods across the entire SNR spectrum. To provide a more granular assessment of structural recovery capability—as requested by the statistical motivation of our MAP framework—we summarize the precision, recall, and F1-scores in Table 1 for the case of $\text{SNR}_Y = 15$ dB.

Table 1. Topology recovery performance comparison under $T = 100$ and $\text{SNR}_Y = 15$ dB.

Method	Precision	Recall	F1-Score	MSE _A
LS-SEM	0.233	0.355	0.281	0.228
TLS-SEM	0.320	0.316	0.318	0.218
sTLS-SEM	0.261	0.303	0.280	0.193
XLS-SEM (Ours)	0.352	0.329	0.340	0.184

While LS-SEM and TLS-variants are limited by the quality of the provided input measurements $\tilde{\mathbf{X}}$, XLS-SEM effectively “denoises” the system by leveraging the low-rank and spatio-temporal structural priors of the latent inputs. As shown in Table 1, although the restricted sample size ($T = 100$) makes joint identification highly challenging, XLS-SEM achieves the highest precision and F1-score. This advantage remains consistent despite the fact that XLS-SEM performs the task in a completely blind manner, whereas the baselines rely on noisy prior observations of the inputs. This confirms that the joint structural priors effectively compensate for the absence of direct measurements, allowing XLS-SEM to

exceed the performance ceiling typically imposed by corrupted input data in conventional SEM frameworks.

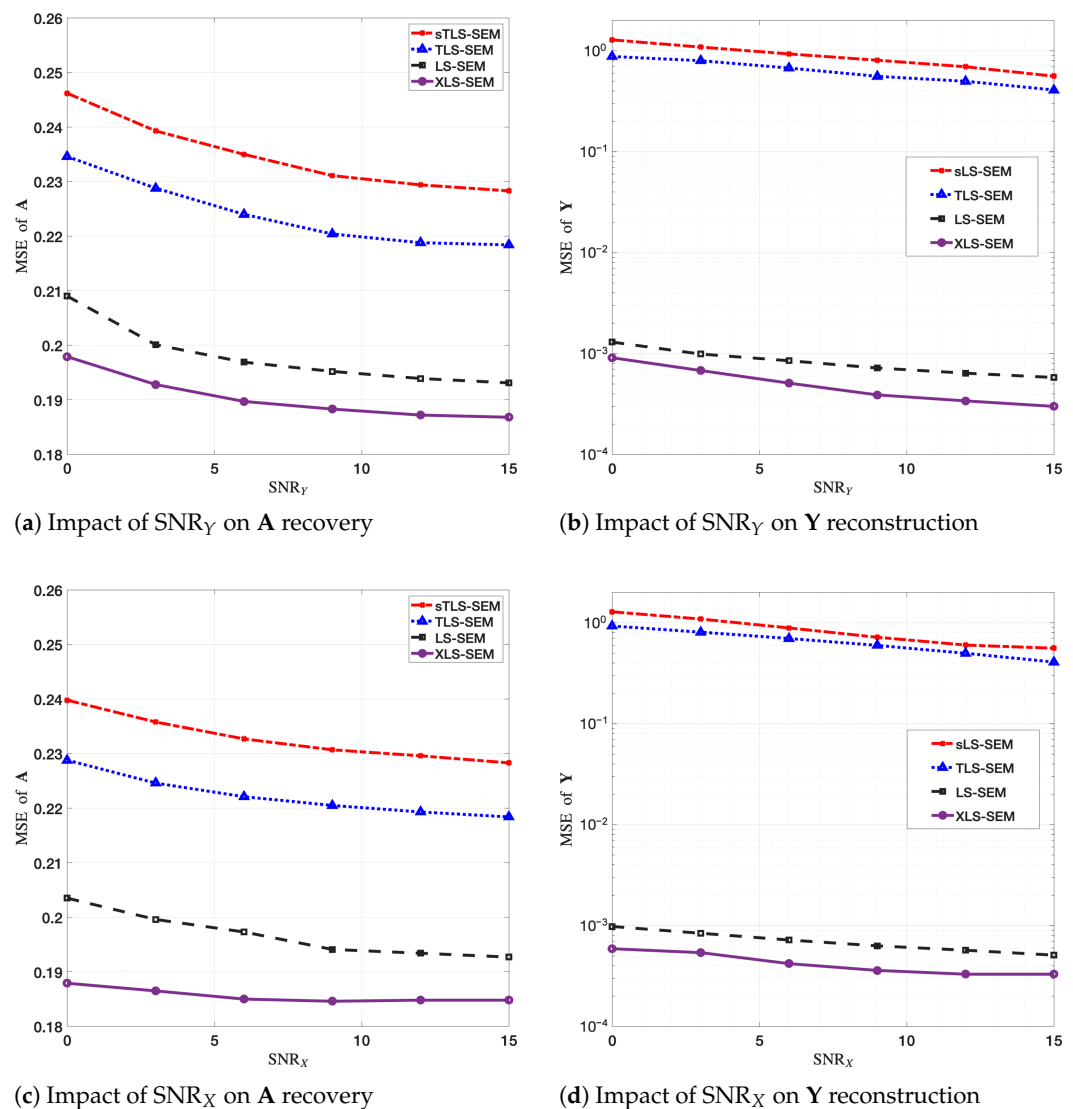


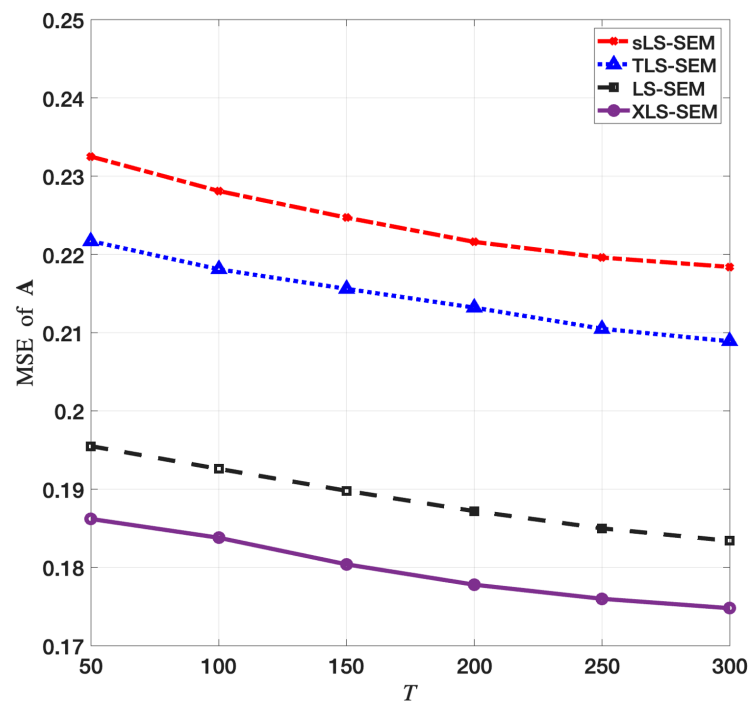
Figure 3. Performance comparison under varying signal-to-noise Ratio (SNR). (a,b) Metrics for varying SNR_Y ; (c,d) metrics for varying SNR_X .

4.1.3. Impact of Sample Size T

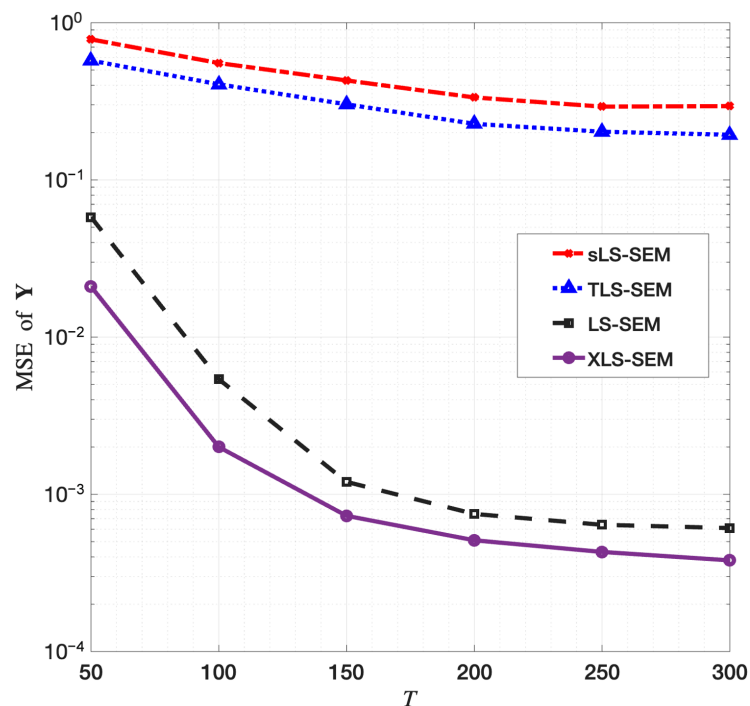
The influence of the temporal observation length T is evaluated with SNR_Y and SNR_X fixed at 15 dB. As shown in Figure 4, an increase in T provides more data points for the ADMM framework to resolve the underlying causal links and latent profiles, leading to a visible reduction in MSE for all methods. XLS-SEM exhibits superior sample efficiency, achieving lower estimation errors even at smaller values of T . For the output reconstruction (MSE_Y), the gap between XLS-SEM and the baselines widens as T increases, suggesting that our model more effectively exploits temporal consistency ($\|XD_T\|_2$) as more temporal information becomes available.

To further investigate the asymptotic identifiability of the proposed framework, we evaluate structural recovery performance under an extended observation window. As the sample size T increases from 100 to 1000, the F1-scores of XLS-SEM exhibit a steady and significant upward trend, rising from 0.340 to 0.482. While the joint identification of A and X remains inherently challenging in a completely blind setting, this performance gain confirms the statistical consistency of our approach. It demonstrates that the incorporated

joint spatio-temporal and low-rank priors allow the algorithm to effectively resolve structural ambiguities and suppress observation noise as more temporal information becomes integrated.



(a) Impact of T on $\mathbf{A}$ recovery



(b) Impact of T on $\mathbf{Y}$ reconstruction

Figure 4. Performance comparison under varying sample size T . (a) Topology estimation error; (b) signal reconstruction error.

4.1.4. Impact of Latent Input Complexity (Q and P)

Beyond signal quality and length, we investigate the impact of the structural complexity of $\mathbf{X}$ by varying the number of input channels Q and the spatial influence per channel P , with $\text{SNR}_Y = 15$ dB and $T = 100$.

When varying Q with $P = 2$ (Figure 5a,b), most baselines maintain a relatively stable yet inferior performance. XLS-SEM consistently maintains the highest accuracy, although its performance shows a slight, gradual degradation as Q increases. This phenomenon is likely due to the fact that a larger number of input channels increases the density of the exogenous matrix $\mathbf{X}$, which marginally weakens the effectiveness of the row-sparsity ($\ell_{2,1}$) and low-rank (nuclear norm) priors. A similar trend is observed when varying the number of influenced nodes P (Figure 5c,d). Despite the increased complexity of the exogenous stimuli, XLS-SEM remains robust, demonstrating that the joint regularization framework can effectively separate dense or multi-channel latent inputs from the underlying graph topology.

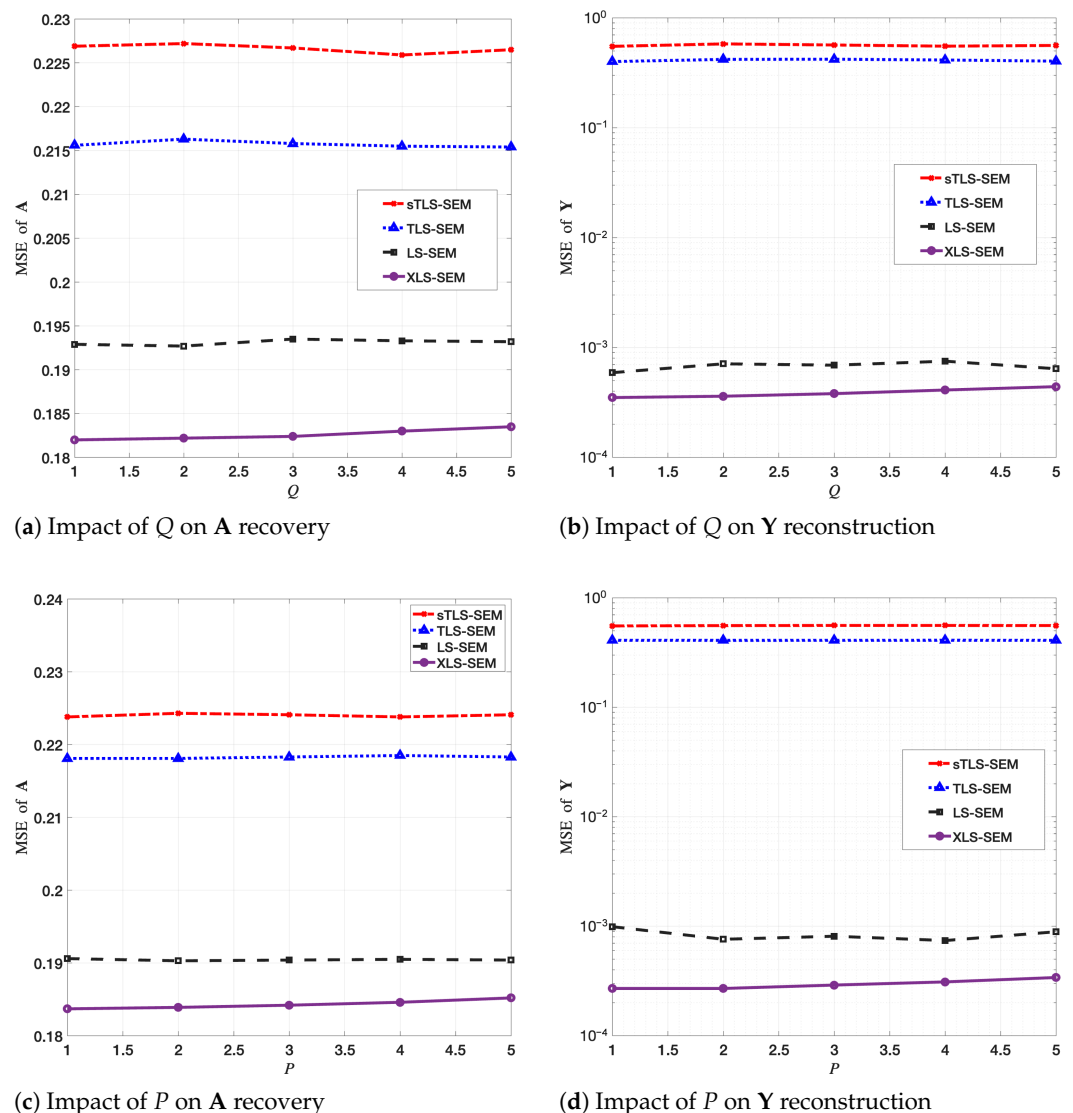


Figure 5. Performance comparison under varying latent input complexity Q or P . (a,b) Performance versus number of channels Q ; (c,d) performance versus spatial influence P .

4.2. Real-World Dataset

4.2.1. Diabetes Clinical Records

In this subsection, we evaluate the proposed framework using a real-world diabetes dataset from the UCI Machine Learning Repository [36], which comprises longitudinal clinical records from 70 patients. Each record provides comprehensive daily profiles of over

ten clinical indicators, with four measurements per day capturing blood glucose levels, insulin dosages, and physical activity schedules.

The model variables are constructed through rigorous preprocessing. The output matrix $\mathbf{Z} \in \mathbb{R}^{70 \times 120}$ represents blood glucose measurements across $N = 70$ patients over $T = 120$ time points. To handle data irregularities, missing values are forward-filled along the temporal dimension, and the entries of $\mathbf{Z}$ are subsequently mapped to the $[0, 1]$ range via min–max normalization.

To accurately reflect the physiological impact of interventions, we incorporate clinically established pharmacokinetic duration windows: regular insulin (5 h), NPH insulin (12 h), ultra-long-acting insulin (27 h), and meals/exercise (≈ 2 h). These effects are modeled as discrete-time diffusion processes by convolving each input at time t with a rectangular window $w(\tau)$ of width $\lceil \text{duration/sampling interval} \rceil$. For instance, the NPH insulin signal is diffused as $\tilde{X}_{\text{NPH}}(t) = \sum_{\tau=-1}^1 w(\tau) \cdot X_{\text{NPH}}(t - \tau)$, while ultra-long-acting insulin spans five time steps. Exercise inputs are assigned a 2-h window, influencing both the current and the subsequent time step.

In this context, $\mathbf{Z}$ tracks blood glucose trajectories, while exogenous factors—including exercise and insulin administration—serve as latent input signals. To account for patient heterogeneity, the graph topology $\mathbf{A}$ is designed to encode similarities among patients within the same diabetes subtype. This allows the model to capture individualized glucose dynamics while identifying shared regulatory mechanisms across the cohort. By applying the same structural priors validated in our synthetic experiments, XLS-SEM identifies a sparse yet physically interpretable interaction network, which reflects the shared physiological responses among patients under similar insulin–meal protocols.

Figure 6 illustrates the normalized fitting loss as a function of the time horizon T , defined as $\|\mathbf{Z} - \mathbf{AZ} - \mathbf{X}\|_F^2 / \|\mathbf{Z}\|_F^2$, where $\mathbf{Z}$ denotes the observation matrix, and $\mathbf{A}, \mathbf{X}$ represent the estimated parameters. As expected, increasing the observation window T provides more data for the model to resolve underlying dependencies, thereby improving reconstruction accuracy. Notably, XLS-SEM consistently achieves the lowest reconstruction loss. Consistent with the MAP-based formulation, the superior alignment with clinical data demonstrates that the joint incorporation of low-rank and spatio-temporal smoothness priors effectively regularizes the identification process, allowing the model to distinguish true physiological dependencies from the measurement noise and sensor missingness inherent in real-world longitudinal records.

Beyond numerical fitting, the graph topology $\mathbf{A}$ identified by XLS-SEM reveals the latent physiological interdependencies within the patient cohort. Specifically, the sparse connections in $\mathbf{A}$ tend to cluster patients who exhibit similar glucose–insulin sensitivity profiles. This suggests that the model effectively captures shared regulatory mechanisms by identifying correlations in glucose fluctuations. By disentangling the individual-specific transient noise through $\mathbf{X}$, the recovered $\mathbf{A}$ provides a robust representation of common metabolic dynamics, which is consistent with the physiological variations expected in type-1 diabetes heterogeneity.

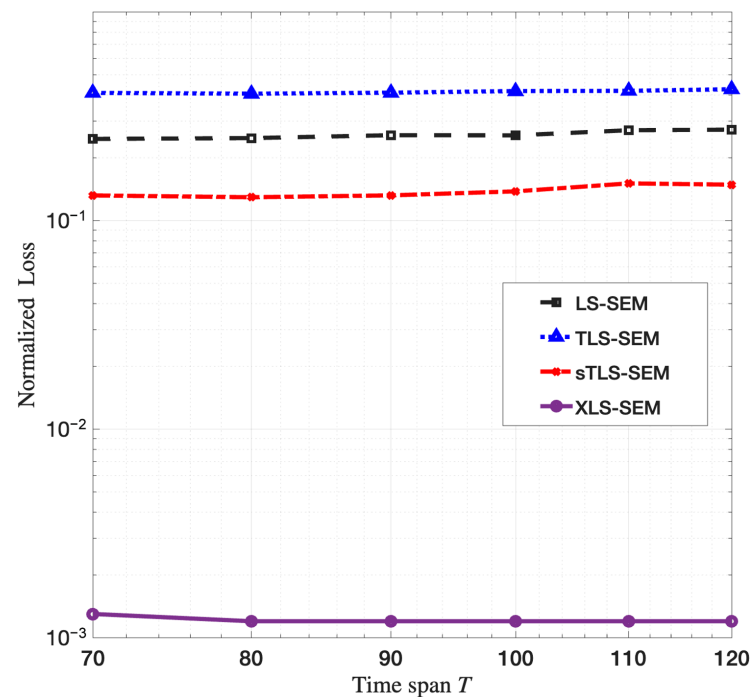


Figure 6. Normalized fitting loss versus observation length T for the Diabetes dataset.

4.2.2. Beijing Air Quality Data

We further evaluate the framework by inferring the topology of an air quality monitoring network in Beijing, focusing on the influence of extreme meteorological events. This study utilizes hourly data from 12 national monitoring stations (UCI Repository). To characterize the overall pollution level, we construct a composite air quality signal $\mathbf{Z}$ by summing the min–max normalized concentrations of $\text{PM}_{2.5}$ and O_3 .

The exogenous input $\mathbf{X}$ represents the aggregate intensity of extreme meteorological conditions. It is computed by summing the normalized values of temperature, dew point, atmospheric pressure, and wind speed, specifically when these variables exceed severity thresholds (0.8 for temperature, dew point, and pressure; 0.7 for wind speed). The spatial relationships inferred by the model are validated against a ground-truth adjacency matrix $\mathbf{A}$ constructed based on the geographical proximity of the stations.

Figure 7 depicts the variation of the normalized fitting loss with respect to the time span T . The results indicate that all algorithms exhibit improved fitting performance as the sample size increases. However, XLS-SEM maintains a consistent performance lead over LS-SEM, TLS-SEM, and sTLS-SEM. The identified spatial relationships inferred by XLS-SEM reveal a topology that closely aligns with the geographical diffusion patterns of pollutants between monitoring stations. This superior alignment with observational data demonstrates that the comprehensive temporal features captured by XLS-SEM allow the model to better account for the delayed and persistent effects typically associated with meteorological impacts on air quality, further validating the structural consistency of our framework in uncovering physical network dependencies without prior knowledge of exact input trajectories.

The structural validity of the inferred topology $\mathbf{A}$ is further confirmed by its alignment with the geographical distribution of the monitoring stations. The dominant edges in $\mathbf{A}$ correspond to stations located along potential pollutant transport paths, such as the North–South axis of the Beijing metropolitan area. This indicates that XLS-SEM successfully identifies the directional diffusion patterns of $\text{PM}_{2.5}$ and O_3 pollutants. Unlike baseline methods that are easily biased by unmeasured local emissions, our framework utilizes the

low-rank prior on $\mathbf{X}$ to isolate these local disturbances, thereby uncovering intrinsic spatial dependencies that are consistent with regional atmospheric transport patterns.

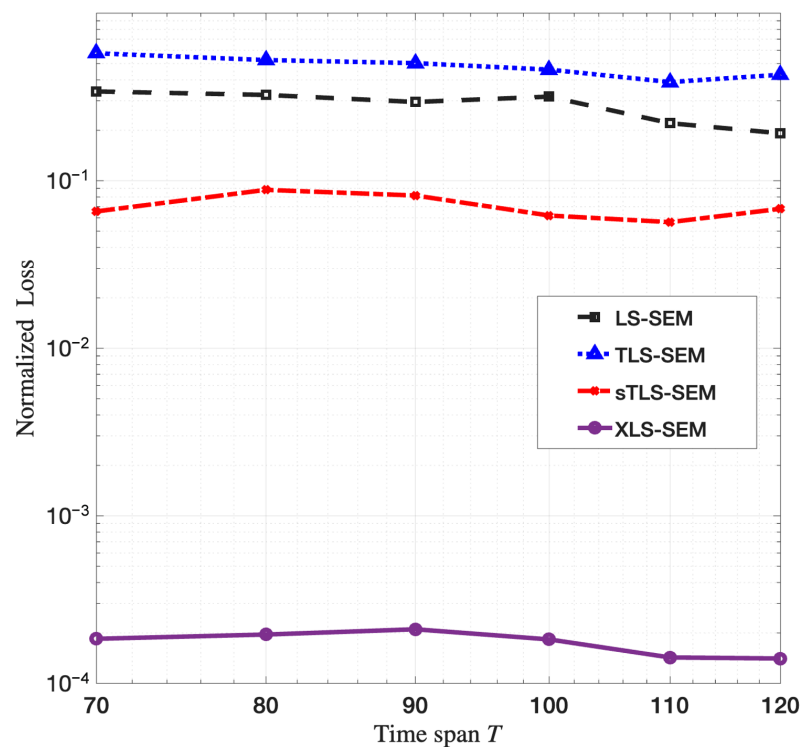


Figure 7. Normalized fitting loss versus observation length T for the Beijing air quality dataset.

5. Conclusions

This paper addresses the challenge of blind network topology identification under latent exogenous inputs. Unlike conventional SEM approaches that often confound external stimuli with endogenous interactions, we proposed a spatio-temporally structured SEM framework that jointly infers graph topology and unobserved driving signals.

The core innovation is the structure-aware modeling of latent inputs, incorporating spatial sparsity and temporal piecewise smoothness. Through the development of an MAP-based joint identification framework, this prior transforms the ill-posed blind identification task into a tractable regularized optimization problem.

Experimental results on synthetic and real-world datasets demonstrate that our method significantly outperforms state-of-the-art sparse SEM baselines. Specifically, the proposed framework achieves superior performance not only in numerical reconstruction accuracy but also in structural recovery metrics, as evidenced by the high F1-scores and the observed asymptotic consistency. By effectively disentangling internal dynamics from latent drivers, the proposed framework provides a robust tool for analyzing complex systems with unmeasured inputs. Future work will extend this model to time-varying topologies and develop automated criteria for determining the number of latent sources.

Author Contributions: Conceptualization, J.Z.; methodology, J.Z.; software, J.Z.; validation, J.Z., R.Y. and X.S.; formal analysis, J.Z.; investigation, J.Z.; resources, J.Z.; data curation, J.Z.; writing—original draft preparation, J.Z.; writing—review and editing, J.Z., R.Y., X.S. and S.F.; visualization, J.Z.; supervision, J.Z.; project administration, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: The authors would like to thank their supervisors and colleagues for their invaluable guidance, technical support, and constructive comments throughout the preparation of this manuscript. The authors also appreciate the anonymous reviewers for their insightful suggestions that helped improve the quality of this work.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

SEM	Structural Equation Modeling
ADMM	Alternating Direction Method of Multipliers
MSE	Mean Squared Error
SNR	Signal-to-Noise Ratio
PDHG	Primal-Dual Hybrid Gradient

References

- Shuman, D.I.; Narang, S.K.; Frossard, P.; Ortega, A.; Vandergheynst, P. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Process. Mag.* **2013**, *30*, 83–98. [\[CrossRef\]](#)
- Leus, G.; Marques, A.G.; Moura, J.M.F.; Ortega, A.; Shuman, D.I. Graph signal processing: History, development, impact, and outlook. *IEEE Signal Process. Mag.* **2023**, *40*, 49–60. [\[CrossRef\]](#)
- Yan, Y.; Hou, J.; Song, Z.; Kuruoglu, E.E. Signal processing over time-varying graphs: A systematic review. *arXiv* **2024**, arXiv:2412.00462. [\[CrossRef\]](#)
- Sandryhaila, A.; Moura, J.M.F. Discrete signal processing on graphs. *IEEE Trans. Signal Process.* **2013**, *61*, 1644–1656. [\[CrossRef\]](#)
- Ortega, A.; Frossard, P.; Kovacevic, J.; Moura, J.M.F.; Vandergheynst, P. Graph Signal Processing: Overview, Challenges, and Applications. *Proc. IEEE* **2018**, *106*, 808–828. [\[CrossRef\]](#)
- Kaplan, D. *Structural Equation Modeling: Foundations and Extensions*, 2nd ed.; Sage Publications: Thousand Oaks, CA, USA, 2009.
- Bentler, P.M.; Weeks, D.G. Linear structural equations with latent variables. *Psychometrika* **1980**, *45*, 289–308. [\[CrossRef\]](#)
- Jöreskog, K.G. A general method for estimating a linear structural equation system. In *Structural Equation Modeling*; Academic Press: New York, NY, USA, 1973; pp. 85–112.
- Friston, K.J. Functional and effective connectivity: A review. *Brain Connect.* **2011**, *1*, 13–36. [\[CrossRef\]](#)
- Giannakis, G.B.; Shen, Y.; Karanikolas, G.V. Monitoring and optimization of cyber-physical networks: A graph signal processing approach. *IEEE Signal Process. Mag.* **2018**, *35*, 34–46.
- Mei, J.; Moura, J.M.F. Signal processing on graphs: Estimating the structure of a graph. *IEEE Trans. Signal Process.* **2017**, *65*, 2045–2058.
- Baingana, B.; Mateos, G.; Giannakis, G.B. Proximal-gradient algorithms for tracking cascades over social networks. *IEEE J. Sel. Top. Signal Process.* **2014**, *8*, 563–575. [\[CrossRef\]](#)
- Cai, X.; Bazerque, J.A.; Giannakis, G.B. Inference of gene regulatory networks with sparse structural equation models exploiting genetic perturbations. *PLoS Comput. Biol.* **2013**, *9*, e1003068. [\[CrossRef\]](#)
- Giannakis, G.B.; Shen, Y.; Karanikolas, G.V. Topology identification and learning over graphs: Accounting for nonlinearities and dynamics. *Proc. IEEE* **2018**, *106*, 787–807. [\[CrossRef\]](#)
- Goldberger, A.S. Structural equation methods in the social sciences. *Econometrica* **1972**, *40*, 979–1001. [\[CrossRef\]](#)
- Muthén, B. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* **1984**, *49*, 115–132. [\[CrossRef\]](#)
- Ning, X.; Karypis, G. SLIM: Sparse linear methods for top-N recommender systems. In Proceedings of the 2011 IEEE 11th International Conference on Data Mining (ICDM), Vancouver, BC, Canada, 11–14 December 2011; pp. 497–506.
- Browne, M.W. Generalized least squares estimators in the analysis of covariance structures. *S. Afr. Stat. J.* **1974**, *8*, 1–24. [\[CrossRef\]](#)
- Bollen, K.A. An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika* **1996**, *61*, 109–121. [\[CrossRef\]](#)

20. Ceci, E.; Shen, Y.; Giannakis, G.B.; Barbarossa, S. Graph-based learning under perturbations via total least-squares. *IEEE Trans. Signal Process.* **2020**, *68*, 2870–2882. [[CrossRef](#)]
21. Shames, I.; Teixeira, A.M.H.; Sandberg, H.; Johansson, K.H. Distributed identification of network topology with noisy node data. In Proceedings of the 51st IEEE Conference on Decision and Control (CDC), Maui, HI, USA, 10–13 December 2012; pp. 5205–5210.
22. Preti, M.G.; Bolton, T.A.W.; Griffa, A.; Van De Ville, D. Graph signal processing for neuroimaging to reveal dynamics of brain structure-function coupling. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.
23. Mazurek, K.; Hohol, M. Revisiting information cascades in online social networks. *Mathematics* **2024**, *13*, 77. [[CrossRef](#)]
24. Segarra, S.; Marques, A.G.; Mateos, G.; Ribeiro, A. Blind identification of graph filters. *IEEE Trans. Signal Process.* **2017**, *65*, 1146–1159. [[CrossRef](#)]
25. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis? *J. ACM* **2011**, *58*, 1–37. [[CrossRef](#)]
26. Thanou, D.; Shuman, D.I.; Frossard, P. Learning parametric dictionaries for graph signals. *IEEE Trans. Signal Process.* **2017**, *65*, 2517–2530.
27. Tibshirani, R.; Saunders, M.; Rosset, S.; Zhu, J. Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 91–108. [[CrossRef](#)]
28. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **2011**, *3*, 1–122. [[CrossRef](#)]
29. Wang, Y.; Liu, J.; Zhang, X. Low-rank matrix recovery via nonconvex optimization and ADMM algorithm. *Mathematics* **2023**, *11*, 652.
30. Hong, M.; Luo, Z.Q.; Razaviyayn, M. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *SIAM J. Optim.* **2016**, *26*, 337–364. [[CrossRef](#)]
31. Liu, Q.; Shen, Z.; Gu, Y. Linearized ADMM for nonconvex nonsmooth optimization with convergence analysis. *IEEE Access* **2019**, *7*, 76131–76144. [[CrossRef](#)]
32. Bollen, K.A. *Structural Equations with Latent Variables*; John Wiley & Sons: New York, NY, USA, 1989.
33. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1996**, *58*, 267–288. [[CrossRef](#)]
34. Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202. [[CrossRef](#)]
35. Chambolle, A.; Pock, T. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **2011**, *40*, 120–145. [[CrossRef](#)]
36. Kahn, M. Diabetes Data Set. UCI Machine Learning Repository. Available online: <https://archive.ics.uci.edu/dataset/34/diabetes> (accessed on 15 February 2026).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.