# A Covariance Matching Approach to Graph Topology Identification

Yongsheng Han, Raj Thilak Rajan, Geert Leus

TU Delft, Signal Processing Systems

GSP Workshop 2026

Extended from the ICASSP 2025 version [1]

Full paper link

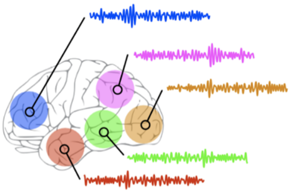
arXiv:2601.15999

[1] Yongsheng Han, Alberto Natali, and Geert Leus, *Graph Topology Identification Based on Covariance Matching*, ICASSP 2025.

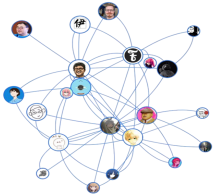
# Graph Topology Identification

## Understanding Hidden Relationships

In many fields, relationships among entities are not directly observable. Graph topology identification helps to infer these hidden structures from nodal data.



Brain Connectivity Network



Social Interaction Network

# Graph Topology Identification (GTI)

## General Process

- GTI relies on the fact that the nodal data is related to the graph.
- Specifically, it follows a distribution determined by the graph

$$\mathbf{x} \sim \mathscr{F}(\mathbf{S}).$$

- For a Gaussian distribution:

$$\mathscr{F}(\mathbf{S}) = \mathcal{N}(\boldsymbol{\mu}(\mathbf{S}), \boldsymbol{\Sigma}(\mathbf{S})).$$

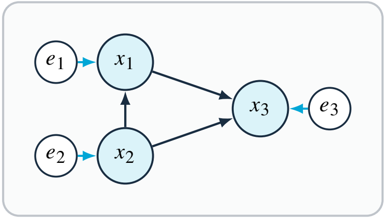
- Here, both $\boldsymbol{\mu}(\mathbf{S})$ and $\boldsymbol{\Sigma}(\mathbf{S})$ are functions of the graph structure $\mathbf{S}$.

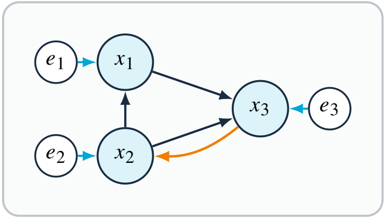
## Common Approaches

| Smoothness [1] | Graphical Lasso [2] | Spectral Template [3] | SEM [4] |
|---|---|---|---|
| graph signals vary smoothly | sparse precision matrix | shared eigenvectors | structural equations |

[1] Kalofolias, *How to Learn a Graph from Smooth Signals*, AISTATS 2016.
[2] Friedman, Hastie, and Tibshirani, *Sparse Inverse Covariance Estimation with the Graphical Lasso*, Biostatistics 2008.
[3] Segarra et al., *Network Topology Inference from Spectral Templates*, IEEE Transactions on Signal and Information Processing over Networks 2017.
[4] Pearl, *Graphs, Causality, and Structural Equation Models*, Sociological Methods & Research 1998.

# Structural Equation Model (SEM)




## Setup

- Each node signal is affected by other nodes:

$$\mathbf{x} = \mathbf{S}\mathbf{x} + \mathbf{e}, \quad \mathbf{x} = (\mathbf{I} - \mathbf{S})^{-1}\mathbf{e}.$$

- $\mathbf{S}$ encodes node-to-node influence, with $\text{diag}(\mathbf{S}) = \mathbf{0}$.
- With Gaussian noise $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$:

$$\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{x}), \quad \boldsymbol{\Sigma}_\mathbf{x} = (\mathbf{I} - \mathbf{S})^{-1}(\mathbf{I} - \mathbf{S})^{-\top}.$$

- **Problem Statement:**
  Observing $T$ realizations $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]$, estimate $\mathbf{S}$.

[1] Pearl, *Graphs, Causality, and Structural Equation Models*, Sociological Methods & Research 1998.
[2] Bollen, *Structural Equations with Latent Variables*, Wiley 1989.

# Linear Regression

## Optimization Problem

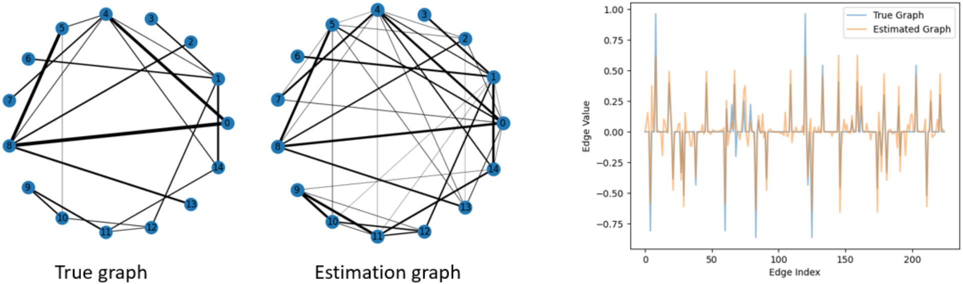
- **SigMatch** seeks $\mathbf{S}$ by minimizing:

$$\min_{\hat{\mathbf{S}}} \ \|\mathbf{X} - \hat{\mathbf{S}}\,\mathbf{X}\|_F^2 \quad \text{subject to} \quad \operatorname{diag}(\hat{\mathbf{S}}) = \mathbf{0}.$$

- **Not** the maximum likelihood estimator in general.
- Only safe in special cases:
  - A directed acyclic graph (DAG) [1]
  - Deterministic known exogenous variables [2]

[1] Peters and Bühlmann, *Identifiability of Gaussian Structural Equation Models with Equal Error Variances*, Biometrika 2014.
[2] Cai, Bazerque, and Giannakis, *Inference of Gene Regulatory Networks with Sparse Structural Equation Models Exploiting Genetic Perturbations*, PLOS Computational Biology 2013.

# Analysis of SigMatch




Visualization of SigMatch outcomes.

## Convergence Issue

For simple undirected graphs with loops, the SigMatch method **fails to converge to the correct graph** even $T \to \infty$.

# Covariance Matching

## Covariance Matching Framework

- **Goal**: Estimate $\mathbf{S}$ so that the theoretical covariance

$$\boldsymbol{\Sigma}_x = (\mathbf{I}-\mathbf{S})^{-1}(\mathbf{I}-\mathbf{S})^{-\top}$$

  is close to the *sample* covariance

$$\mathbf{C}_\mathbf{x} = \frac{1}{T}\mathbf{X}\mathbf{X}^\top.$$

- **Why covariance?**
  - For zero-mean Gaussian data, covariance determines the distribution.
  - Closer to maximum likelihood estimator.

# Swapping the Objective and the Constraint

- Classic covariance fitting between variable $\hat{\boldsymbol{\Sigma}}_x$ (depends on variable $\hat{\mathbf{S}}$) and $\mathbf{C}_\mathbf{x}$.

**Classical covariance fitting**

$$\begin{aligned} \min_{\hat{\mathbf{S}}} \quad & \|\hat{\boldsymbol{\Sigma}}_\mathbf{x} - \mathbf{C}_\mathbf{x}\|_F^2 \\ \text{s.t.} \quad & \hat{\mathbf{S}} \in \mathcal{S} \end{aligned}$$

# Swapping the Objective and the Constraint

- Classic covariance fitting between variable $\hat{\boldsymbol{\Sigma}}_x$ (depends on variable $\hat{\mathbf{S}}$) and $\mathbf{C}_\mathbf{x}$.
- So first make $\hat{\boldsymbol{\Sigma}}_\mathbf{x}$ match $\mathbf{C}_\mathbf{x}$.

**Classical covariance fitting**

$$\begin{aligned} \min_{\hat{\mathbf{S}}} \quad & \left\|\hat{\boldsymbol{\Sigma}}_\mathbf{x} - \mathbf{C}_\mathbf{x}\right\|_F^2 \\ \text{s.t.} \quad & \hat{\mathbf{S}} \in \mathcal{S} \end{aligned}$$

$\xrightarrow{\text{swap fit/ constraint}}$

**CovMatch**

$$\begin{aligned} \min_{\hat{\mathbf{S}}} \quad & r_{\mathcal{S}}(\hat{\mathbf{S}}) \\ \text{s.t.} \quad & \hat{\boldsymbol{\Sigma}}_\mathbf{x} = \mathbf{C}_\mathbf{x} \end{aligned}$$

# Swapping the Objective and the Constraint

- Classic covariance fitting between variable $\hat{\mathbf{\Sigma}}_x$ (depends on variable $\hat{\mathbf{S}}$) and $\mathbf{C}_\mathbf{x}$.
- So first make $\hat{\mathbf{\Sigma}}_\mathbf{x}$ match $\mathbf{C}_\mathbf{x}$.
- Then use hollowness and sparsity (or any structure) to pick a particular graph.

**Classical covariance fitting**

$$\min_{\hat{\mathbf{S}}} \quad \|\hat{\mathbf{\Sigma}}_\mathbf{x} - \mathbf{C}_\mathbf{x}\|_F^2 \qquad \text{s.t.} \quad \hat{\mathbf{S}} \in \mathcal{S}$$

$\xrightarrow{\text{swap fit/ constraint}}$

**CovMatch**

$$\min_{\hat{\mathbf{S}}} \quad r_{\mathcal{S}}(\hat{\mathbf{S}}) \qquad \text{s.t.} \quad \hat{\mathbf{\Sigma}}_\mathbf{x} = \mathbf{C}_\mathbf{x}$$

**Fit covariance exactly. Then use structure to choose a graph.**

# How to Match Covariance

- **How do we make $\hat{\Sigma}_x = \mathbf{C}_x$**

# How to Match Covariance

- **How do we make $\hat{\boldsymbol{\Sigma}}_{\mathrm{x}} = \mathbf{C}_{\mathbf{x}}$**

**Undirected CovMatch [1]**

$\mathbf{C}_{\mathbf{x}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t^{\top}$

$\hat{\boldsymbol{\Sigma}}_{\mathrm{x}} = (\mathbf{I} - \hat{\mathbf{S}})^{-2}$

[1] Yongsheng Han, Alberto Natali, and Geert Leus, *Graph Topology Identification Based on Covariance Matching*, ICASSP 2025.

# How to Match Covariance

- **How do we make $\hat{\boldsymbol{\Sigma}}_{\mathrm{x}} = \mathbf{C}_{\mathbf{x}}$**

**Undirected CovMatch [1]**

$$\mathbf{C}_{\mathbf{x}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t^\top \xrightarrow{\text{EVD}} \mathbf{C}_{\mathbf{x}} = \mathbf{U}_{\mathbf{x}}\,\mathrm{diag}(\boldsymbol{\lambda}_{\mathbf{x}})\mathbf{U}_{\mathbf{x}}^\top$$

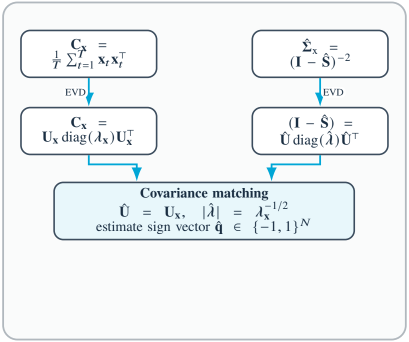
$$\hat{\boldsymbol{\Sigma}}_{\mathrm{x}} = (\mathbf{I} - \hat{\mathbf{S}})^{-2} \xrightarrow{\text{EVD}} (\mathbf{I} - \hat{\mathbf{S}}) = \hat{\mathbf{U}}\,\mathrm{diag}(\hat{\boldsymbol{\lambda}})\hat{\mathbf{U}}^\top$$

[1] Yongsheng Han, Alberto Natali, and Geert Leus, *Graph Topology Identification Based on Covariance Matching*, ICASSP 2025.

# How to Match Covariance

- **How do we make $\hat{\boldsymbol{\Sigma}}_{\mathrm{x}} = \mathbf{C}_{\mathbf{x}}$**

**Undirected CovMatch [1]**

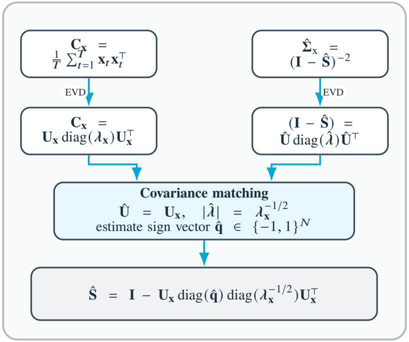



[1] Yongsheng Han, Alberto Natali, and Geert Leus, *Graph Topology Identification Based on Covariance Matching*, ICASSP 2025.

# How to Match Covariance

- **How do we make $\hat{\boldsymbol{\Sigma}}_{\mathrm{x}} = \mathbf{C}_{\mathbf{x}}$**

**Undirected CovMatch [1]**




**Undirected graph optimization problem**

$$\min_{\hat{\mathbf{q}}\in\{-1,1\}^N} \|\,\mathrm{diag}(\hat{\mathbf{S}})\,\|_2^2 + \alpha\|\hat{\mathbf{S}}\|_1$$

$$\text{s.t.} \quad \hat{\mathbf{S}} = \mathbf{I} - \mathbf{U}_{\mathbf{x}}\,\mathrm{diag}(\hat{\mathbf{q}})\,\mathrm{diag}(\boldsymbol{\lambda}_{\mathbf{x}}^{-1/2})\mathbf{U}_{\mathbf{x}}^{\top}$$

- Only the sign ambiguity is left
- We can solve that with off-the-shelf mixed-integer or conic solvers

[1] Yongsheng Han, Alberto Natali, and Geert Leus, *Graph Topology Identification Based on Covariance Matching*, ICASSP 2025.

# How to Match Covariance

- **How do we make $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}} = \mathbf{C}_{\mathbf{x}}$**

**Directed CovMatch**

$\mathbf{C}_{\mathbf{x}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^\top$

EVD

$\mathbf{C}_{\mathbf{x}} = \mathbf{U}_{\mathbf{x}} \operatorname{diag}(\boldsymbol{\lambda}_{\mathbf{x}}) \mathbf{U}_{\mathbf{x}}^\top$

$\hat{\boldsymbol{\Sigma}}_{\mathbf{x}} = (\mathbf{I} - \hat{\mathbf{S}})^{-1} (\mathbf{I} - \hat{\mathbf{S}})^{-\top}$

SVD

$(\mathbf{I} - \hat{\mathbf{S}}) = \hat{\mathbf{V}} \operatorname{diag}(\hat{\boldsymbol{\lambda}}) \hat{\mathbf{U}}^\top$

**Covariance matching**

$\hat{\mathbf{U}} = \mathbf{U}_{\mathbf{x}}, \quad \hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}_{\mathbf{x}}^{-1/2}$

estimate orthogonal factor $\hat{\mathbf{V}} \in \mathcal{O}(N)$

$\hat{\mathbf{S}} = \mathbf{I} - \hat{\mathbf{V}} \operatorname{diag}(\boldsymbol{\lambda}_{\mathbf{x}}^{-1/2}) \mathbf{U}_{\mathbf{x}}^\top$

**Directed graph optimization problem**

$$\min_{\hat{\mathbf{V}} \in \mathcal{O}(N)} \quad \| \operatorname{diag}(\hat{\mathbf{S}}) \|_2^2 + \alpha \|\hat{\mathbf{S}}\|_1$$

$$\text{s.t.} \quad \hat{\mathbf{S}} = \mathbf{I} - \hat{\mathbf{V}} \operatorname{diag}(\boldsymbol{\lambda}_{\mathbf{x}}^{-1/2}) \mathbf{U}_{\mathbf{x}}^\top$$

- The sign ambiguity becomes a rotation ambiguity
- We solve that with a Riemannian manifold-optimization algorithm

# Optimization over the Orthogonal Group

## From graph structure to an objective in $\hat{\mathbf{V}}$

$$\begin{aligned} \min_{\hat{\mathbf{V}} \in \mathcal{O}(N)} \quad & \| \operatorname{diag}(\hat{\mathbf{S}}) \|_2^2 + \alpha \|\hat{\mathbf{S}}\|_1 \\ \text{s.t.} \quad & \hat{\mathbf{S}} = \mathbf{I} - \hat{\mathbf{V}} \operatorname{diag}(\lambda_{\mathbf{x}}^{-1/2}) \mathbf{U}_{\mathbf{x}}^\top . \end{aligned}$$

[1] Absil, Mahony, and Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press 2008.

# Optimization over the Orthogonal Group

## From graph structure to an objective in $\hat{\mathbf{V}}$

$$\min_{\hat{\mathbf{V}} \in \mathcal{O}(N)} \quad \| \operatorname{diag}(\hat{\mathbf{S}}) \|_2^2 + \alpha \|\hat{\mathbf{S}}\|_1$$
$$\text{s.t.} \quad \hat{\mathbf{S}} = \mathbf{I} - \hat{\mathbf{V}} \operatorname{diag}(\boldsymbol{\lambda}_{\mathbf{x}}^{-1/2}) \mathbf{U}_{\mathbf{x}}^\top.$$

Substituting the definition of $\hat{\mathbf{S}}$ gives

$$\mathcal{J}(\hat{\mathbf{V}}) = \left\| \operatorname{diag}\left( \mathbf{I} - \hat{\mathbf{V}} \operatorname{diag}(\boldsymbol{\lambda}_{\mathbf{x}}^{-1/2}) \mathbf{U}_{\mathbf{x}}^\top \right) \right\|_2^2 + \alpha \left\| \mathbf{I} - \hat{\mathbf{V}} \operatorname{diag}(\boldsymbol{\lambda}_{\mathbf{x}}^{-1/2}) \mathbf{U}_{\mathbf{x}}^\top \right\|_1.$$

Now the only variable is the orthogonal factor $\hat{\mathbf{V}} \in \mathcal{O}(N)$.

[1] Absil, Mahony, and Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press 2008.

# Riemannian Gradient Descent

## Algorithmic Outline

1: **Input:** random orthogonal matrix $\hat{\mathbf{V}}_0 \in \mathcal{O}(N)$
2: **for** $r = 0$ **to** $R-1$ **do**
3: $\quad \boldsymbol{\Gamma}_r \leftarrow \partial \mathcal{J} / \partial \hat{\mathbf{V}}(\hat{\mathbf{V}}_r)$
4: $\quad \mathbf{G}_r \leftarrow \boldsymbol{\Gamma}_r \hat{\mathbf{V}}_r^\top - \hat{\mathbf{V}}_r \boldsymbol{\Gamma}_r^\top$
5: $\quad \mathbf{P}_r \leftarrow \exp(-\mu_r \mathbf{G}_r / \|\mathbf{G}_r\|_F)$
6: $\quad \hat{\mathbf{V}}_{r+1} \leftarrow \mathbf{P}_r \hat{\mathbf{V}}_r$
7: **end for**
8: **return** $\hat{\mathbf{V}}_R$

- The exponential map keeps $\hat{\mathbf{V}}$ orthogonal.

[1] Absil, Mahony, and Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press 2008.

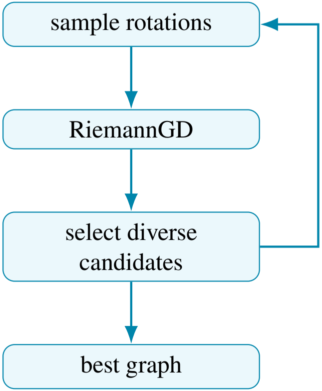
# Escaping Local Minima

### Why basin hopping?

- RiemannGD is only a local search.
- $O(N)$ is too large for naive restarts.

### Basin-hopping strategy

- Perturb candidates by moderate rotations.
- Refine each perturbation by RiemannGD.
- Keep low-cost and diverse candidates.




[1] Wales and Doye, *Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters*, Journal of Physical Chemistry A 1997.
[2] Absil, Mahony, and Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press 2008.

# Experimental Roadmap

## Synthetic Benchmarks

- Graph sizes: $N \in \{20, 40, 60, 80, 100\}$.
- Sample regimes: $T = 1000$ and $T \to \infty$.
- For each $N$, results are averaged over 10 graphs.

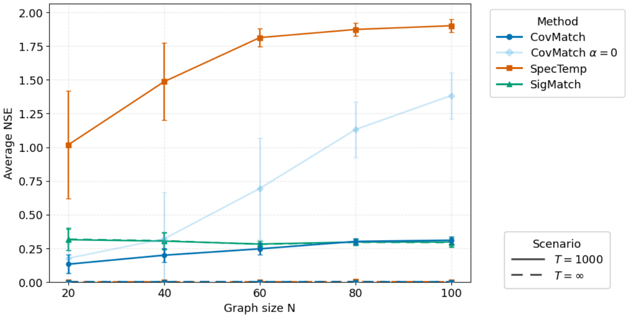
## Graph Models

- Undirected graph with $M = 2N$ edges.
- DAGs with $M = 2N$ expected edges.
- Cyclic directed graphs with $M = 2N$ random directed edges.

## Metric

Average normalized squared error (NSE) between true and estimated adjacency matrices.

# Undirected SEM: CovMatch Scales with Graph Size



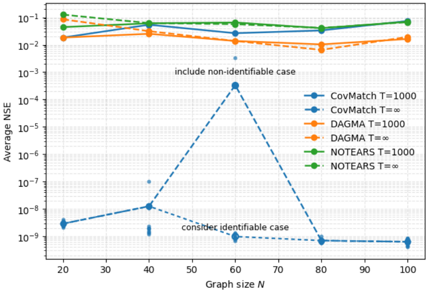

Average NSE versus number of nodes.

## Observations

- Benchmarks:
  - SigMatch [1]
  - SpecTemp [2]
  - CovMatch with $\alpha = 0$ [3]
- As $T \to \infty$, CovMatch achieves nearly zero error.
- Sparsity avoids overly dense finite-sample estimates.
- We prove identifiability for a broad class of undirected graphs.

[1] Aragam, Amini, and Zhou, *Learning Directed Acyclic Graphs with Penalized Neighbourhood Regression*, arXiv:1511.08963, 2015.
[2] Segarra et al., *Network Topology Inference from Spectral Templates*, IEEE Transactions on Signal and Information Processing over Networks 2017.
[3] Han, Natali, and Leus, *Graph Topology Identification Based on Covariance Matching*, ICASSP 2025.

# DAG Benchmark: No Acyclicity Prior Needed



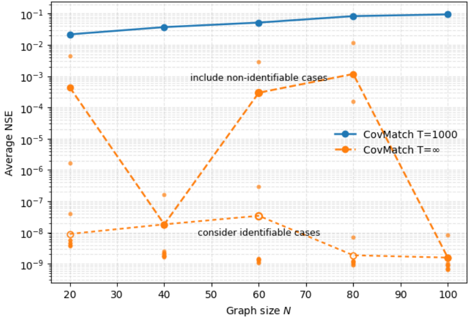

DAG benchmark, average NSE in log scale.

## Observations

- Benchmarks:
  - NOTEARS [1]
  - DAGMA [2]
- We add no acyclicity prior.
- As $T \to \infty$, CovMatch reaches nearly zero error for identifiable cases.
- At $T = 1000$, small graph sizes are comparable to DAGMA [2].
- For larger graph sizes, the error is comparable to NOTEARS [1].


[1] Zheng et al., *DAGs with NO TEARS: Continuous Optimization for Structure Learning*, NeurIPS 2018.
[2] Bello, Aragam, and Ravikumar, *DAGMA: Learning DAGs via M-matrices and a Log-Determinant Acyclicity Characterization*, NeurIPS 2022.

# Cyclic Directed Graphs: Beyond DAGs




Cyclic directed graphs, average NSE in log scale.

## Observations

- CovMatch generalizes to cyclic directed graphs.
- Identifiable cases reach near-zero error.

[1] Bazerque, Baingana, and Giannakis, *Identifiability of Sparse Structural Equation Models for Directed and Cyclic Networks*, GlobalSIP 2013.

# Results and Limitations

## Main Findings

- CovMatch reaches near-zero NSE on sparse identifiable SEMs.
- We prove identifiability for a broad class of undirected graphs.
- On DAGs, it is asymptotically accurate without enforcing acyclicity.
- On cyclic directed graphs, it still recovers identifiable topologies.

## Current Limitations

- We need enough data to compute a good sample covariance matrix.
- The noise covariance needs to be known or reliably estimated.