Partial Homoscedasticity in Causal Discovery with Linear Models

Jun Wu[1], Mathias Drton[1]

[1]Technical University of Munich, Germany;
TUM School of Computation, Information and Technology,
Department of Mathematics and Munich Data Science Institute

ETH-UCPH-TUM Workshop on Graphical Models, October 2022

Outline

Outline

Linear Structural Equation Models

- Linear SEM: random vector X solves

$$X = \Lambda^T X + \epsilon, \qquad \mathrm{Var}[\epsilon] = \mathrm{diag}(\omega).$$

- Represented by a directed aylic graph (DAG): $G = (V, E)$
- The DAG / SEM has a covariance parameterization:

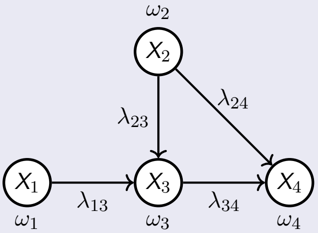
$$\begin{aligned} \phi_G : \mathbb{R}^E \times \mathbb{R}^V &\mapsto PD, \\ (\Lambda, \omega) &\mapsto (I - \Lambda)^{-T} \mathrm{diag}(\omega)(I - \Lambda)^{-1}, \end{aligned}$$

with

$$\mathbb{R}^E = \{ \Lambda \in \mathbb{R}^{V \times V} : \lambda_{ij} = 0 \text{ if } i \to j \notin E \}.$$

Example 1

$$
\begin{aligned}
X_1 &= \varepsilon_1 \\
X_2 &= \varepsilon_2 \\
X_3 &= \lambda_{13} X_1 + \lambda_{23} X_2 + \varepsilon_3 \\
X_4 &= \lambda_{24} X_2 + \lambda_{34} X_3 + \varepsilon_4
\end{aligned}
$$



$$
\Lambda = \begin{pmatrix} 0 & 0 & \lambda_{13} & 0 \\ 0 & 0 & \lambda_{23} & \lambda_{24} \\ 0 & 0 & 0 & \lambda_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3, \omega_4)^T.
$$

Outline

Structural Identifiability

- Classic case: Markov equivalence classes of DAGs
- Homoscedastic errors: if all error variances are equal, the DAG G is uniquely identifiable
  [Chen, Drton, and Wang 2019; Peters and Bühlmann 2014]
- Homoscedastic error, general directed graphs?
- Setup in between: partial homoscedasticity (groupwise equal error variances)

Our contributions

- Derive an implicit description of linear Gaussian SEM under partial homoscedasticity
- Characterize the DAGs that define the same partially homoscedastic linear Gaussian SEM
- Give an algorithm for constructing equivalence class representation

Outline

Partial Homoscedasticity

- A partition of nodes w.r.t. different error variances $\omega_1, \ldots, \omega_K$: $\Pi = \{\pi_1, \ldots, \pi_K\}$. The k'th **(homoscedastic errors) block of nodes** is the set $\pi_k = \{i_1, i_2, \ldots, i_m\}$ such that

$$\omega_{i_1 i_1} = \omega_{i_2 i_2} = \cdots = \omega_{i_m i_m} = \omega_k.$$

- For i, j in the same block: $i \sim_\Pi j$.
- The **groupwise homoscedastic linear Gaussian model** given by $G = (V, E)$ and Π:

$$M_{G,\Pi} = \left\{ \Sigma : \Sigma = \phi_G(\Lambda, \boldsymbol{\omega}),\ \Lambda \in \mathbb{R}^E \text{ and } \boldsymbol{\omega} \in \mathbb{R}^V, \omega_{ii} = \omega_{jj} \text{ if } i \sim_\Pi j \right\}$$

Partial Homoscedasticity

Theorem 1 (Drton 2018)

The error variance ω_{ii} can be computed from the true covariance matrix

$$\omega_{ii} = \sigma_{ii} - \Sigma_{i,A}(\Sigma_{A,A})^{-1}\Sigma_{A,i},$$

for any A such that $\mathrm{pa}(i) \subseteq A \subseteq V \setminus \mathrm{de}(i)$.

- As a corollary, if i and j are in the same block then

$$\sigma_{ii} - \Sigma_{i,A_i}(\Sigma_{A_i,A_i})^{-1}\Sigma_{A_i,i} = \sigma_{jj} - \Sigma_{j,A_j}(\Sigma_{A_j,A_j})^{-1}\Sigma_{A_j,j}$$

holds for all A_i and A_j such that $\mathrm{pa}(i) \subseteq A_i \subseteq V \setminus \mathrm{de}(i)$ and $\mathrm{pa}(j) \subseteq A_j \subseteq V \setminus \mathrm{de}(j)$.

Partial Homoscedasticity

Theorem 2

If i and j are in the same block of partition, then the constraint equation

$$\sigma_{ii} - \Sigma_{i,A_i}(\Sigma_{A_i,A_i})^{-1}\Sigma_{A_i,i} = \sigma_{jj} - \Sigma_{j,A_j}(\Sigma_{A_j,A_j})^{-1}\Sigma_{A_j,j}$$

holds for all matrices $\Sigma \in M_{G,\Pi}$ **if and only if**

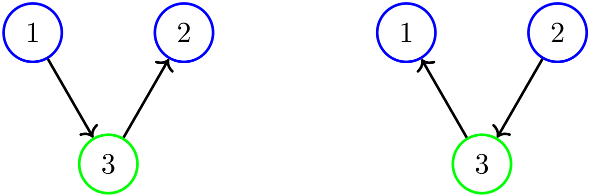
$$\mathrm{pa}(i) \subseteq A_i \subseteq V \setminus \mathrm{de}(i) \text{ and } \mathrm{pa}(j) \subseteq A_j \subseteq V \setminus \mathrm{de}(j).$$

The partially homoscedastic linear Gaussian model $M_{G,\Pi}$ is uniquely determined by

- Equal variance constraints induced by Π
- Conditional independence constraints induced by G

Partial Homoscedasticity

Example: $\Pi = \{\{1,2\},\{3\}\}$



(i) G_1 (ii) G_2

- $\mathcal{M}_1 = \{\Sigma \mid \underbrace{\sigma_{11}\sigma_{33} = \sigma_{22}\sigma_{33} - \sigma_{23}^2}_{\mathrm{Var}[X_2|X_3] = \mathrm{Var}[X_1]},\ \underbrace{\sigma_{13}\sigma_{23} - \sigma_{12}\sigma_{33} = 0}_{X_1 \perp\!\!\!\perp X_2 | X_3},\ \Sigma \in PD_3\}$
- $\mathcal{M}_2 = \{\Sigma \mid \underbrace{\sigma_{22}\sigma_{33} = \sigma_{11}\sigma_{33} - \sigma_{13}^2}_{\mathrm{Var}[X_1|X_3] = \mathrm{Var}[X_2]},\ \underbrace{\sigma_{13}\sigma_{23} - \sigma_{12}\sigma_{33} = 0}_{X_1 \perp\!\!\!\perp X_2 | X_3},\ \Sigma \in PD_3\}$

Outline

Equivalence Classes

Theorem 3

Given a partition Π of nodes, two DAGs G_1, G_2 have the same partially homoscedastic linear Gaussian model ($M_{G_1,\Pi} = M_{G_2,\Pi}$) **if and only if** the following two conditions hold:

1. G_1 and G_2 have the same skeleton and unshielded colliders.
2. For each node i with $i \in \pi_k, |\pi_k| \geq 2$, the parents of i in G_1 and G_2 are the same: $\mathrm{pa}_1(i) = \mathrm{pa}_2(i)$.

In this case we say that G_1 and G_2 are (distributionally) equivalent given the partition Π: $G_1 \approx_\Pi G_2$.

Two extreme cases:

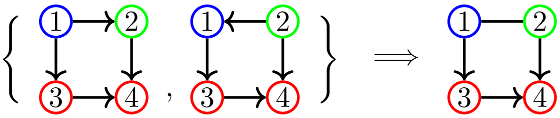
- $\Pi = \Pi_{\text{finest}} = \{\{i\} : i \in V\}$: classic setup, all different variances
- $\Pi = \Pi_{\text{coarsest}} = \{V\}$: all equal variance, uniquely identifiable

Equivalence Classes

Definition: CPDAG

A **completed partially directed acyclic graph** (CPDAG) of a DAG G **under partition** Π is defined as the union of all equivalent DAGs of G:

$$G^*_\Pi := \cup \left(G' \mid G' \approx_\Pi G\right).$$



- Just the classic CPDAG definition, but the construction is a bit different!

Equivalence Classes

The Orientation rules R1-R4 in [Meek 1995], for propagating orientations

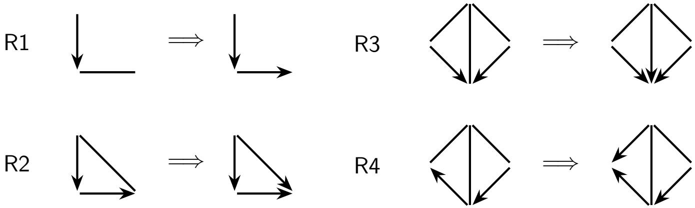


Figure: The 4 Orientation rules.

Equivalence Classes

Algorithm Constructing the equivalence class of a DAG, given the partition

Require: A DAG G, the partition Π

Create an empty graph G'

Copy the skeleton and all edge orientations with unshielded colliders of G to G'

Apply rules R1, R2 and R3 on G' until no more edges can be oriented

for $i \in V$ with $i \in \pi_k$ and $|\pi_k| \geq 2$ **do**

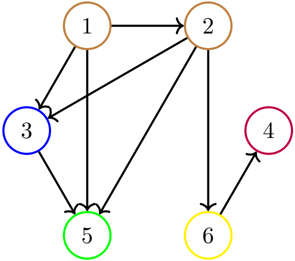
    Copy the orientation of edges in G having one endpoint at i to G'

end for

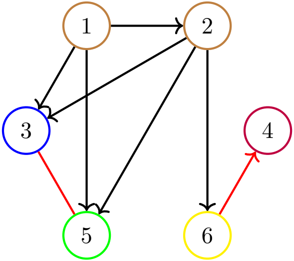
Apply rules R1 and R2 on G' until no more edges can be oriented

return $G' = G^*_\Pi$

Equivalence Classes



(i) DAG G



(ii) CPDAG G'

Figure: A DAG and the corresponding CPDAG, under a fixed partition.

Conclusions

🙂

- Extension of equal variance assumption
- New perspective of model identifibility

🙃

- Restrictive conditions
- Cannot be applied to cycles

References

- Chen, Wenyu, Mathias Drton, and Y Samuel Wang (Sept. 2019). "On causal discovery with an equal-variance assumption". In: *Biometrika* 106.4, pp. 973–980.
- Drton, Mathias (2018). "Algebraic problems in structural equation modeling". In: *The 50th anniversary of Gröbner bases*. Vol. 77. Adv. Stud. Pure Math. Math. Soc. Japan, Tokyo, pp. 35–86.
- Meek, Christopher (1995). "Causal inference and causal explanation with background knowledge". In: *UAI '95: Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence, Montreal, Quebec, Canada, August 18-20, 1995*. Morgan Kaufmann, pp. 403–410.
- Peters, Jonas and Peter Bühlmann (2014). "Identifiability of Gaussian structural equation models with equal error variances". In: *Biometrika* 101.1, pp. 219–228.