Structure Learning of Linear Gaussian Structural Equation Models with Weak Edges

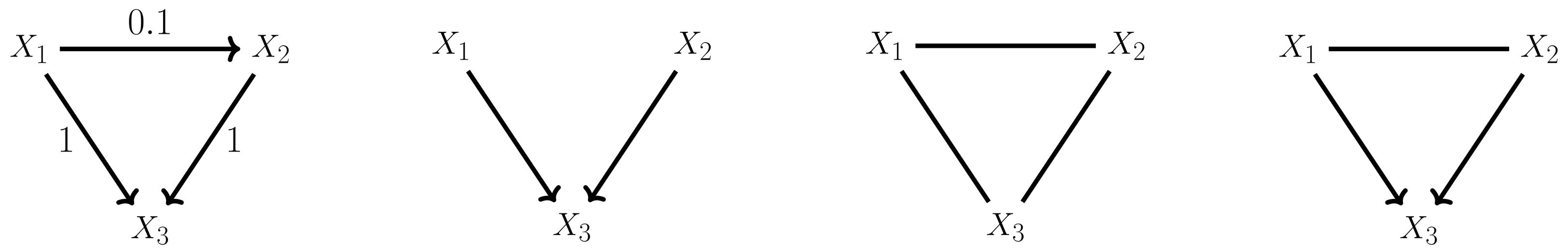
Marco F. Eigenmann[1], Preetam Nandy[2], and Marloes H. Maathuis[1]

[1]Seminar for Statistics, ETH Zurich [2]Department of Biostatistics and Epidemiology, University of Pennsylvania

ETH zürich

Problem: Learning linear Gaussian structural equation models from observational data.

- **Issue:** For increasing sample sizes the output of well known algorithm like GES [1] and PC [3] can become less informative.



True DAG — Smaller sample size — Bigger sample size — Desired result

- **Our contribution:** A new graphical object that contains more orientations than a CPDAG, an algorithm to learn it, theoretical guarantees.

APDAG: An aggregated PDAG (APDAG) is a maximally oriented graph as studied in [2] and is the object we want to learn. It is constructed as follows:

1. Given a multivariate density f of X, compute the solution path of the oracle version of GES [1] for $\lambda \geqslant 0$ (the penalty parameter) and keep the outputs whose skeletons are contained in the skeleton of $C_0 = \text{GES}_0(f)$, the true CPDAG. This yields a set of CPDAGs $\mathcal{C} = \{C_0, \cdots, C_k\}$ with associated penalty parameters $\lambda_0 < \ldots < \lambda_k$.
2. Construct the set of DAGs $\mathcal{G} = \{G_0, \ldots, G_k\}$ consisting of G_0 (the true underlying DAG) restricted to the skeletons of the CPDAGs in $\mathcal{C}$.
3. Construct the CPDAGs $\tilde{\mathcal{C}} = \{\tilde{C}_0, \ldots, \tilde{C}_k\}$ where $\tilde{C}_i = \text{CPDAG}(G_i)$, $0 \leqslant i \leqslant k$.
4. Let $A_0 = \text{AggregateCPDAGs}(\tilde{\mathcal{C}})$ (lines 3 - 15 of Algorithm 1).

Note The APDAG A_0 has the same skeleton as C_0 and

$$\text{DirPart}(C_0) \subseteq \text{DirPart}(A_0) \subseteq \text{DirPart}(G_0).$$

Theoretical results: Soundness and consistency.

Definition (δ-strong faithful) A multivariate Gaussian distribution is said to be *δ-strong faithful* to a DAG $G = (X, E)$ if for every $X_i, X_j \in X$ and for every $S \subseteq X \setminus \{X_i, X_j\}$ it holds that $X_i \not\perp_G X_j | S \Rightarrow |\rho_{X_i, X_j|S}| > \delta$, where $\rho_{X_i, X_j|S}$ is the partial correlation between X_i and X_j given S.

Theorem Given a multivariate Gaussian distribution of X with a perfect map $G_0 = (X, E)$, let $\mathcal{G}$ be the set of DAGs constructed above in step 2, and let $\tilde{\mathcal{C}}$ be the corresponding set of CPDAGs of step 3. Assume that for all $1 \leqslant i \leqslant k$ the distribution of X is δ_i-strong faithful with respect to $G_i \in \mathcal{G}$, where δ_i is such that $\lambda_i = -1/2 \log(1 - \delta_i^2)$. Then $\text{GES}_{\lambda_i}(f) = C_i = \tilde{C}_i$ for all $1 \leqslant i \leqslant k$, and the oracle version of AGES returns the APDAG A_0.

Theorem Under the same conditions of the above theorem, we have

$$\lim_{n \to \infty} \mathbb{P}\left(\text{AGES}(X^{(n)}) = A_0\right) \to 1.$$

Note

- The δ-strong faithfulness assumption is sufficient but not necessary.
- We propose AGES δ-strong faithfulness. It is sufficient and necessary to learn the $\tilde{C}_i$s constructed above.
- Drawback: It is defined through the oracle version of the GES algorithm (see the middle box on the right). Therefore, it cannot be written explicitly and its interpretability is more limited.

The AGES algorithm (sample version)

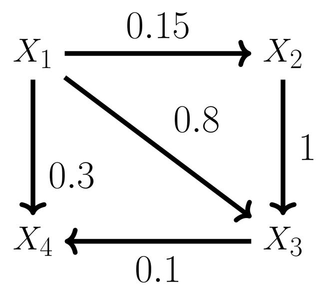
Algorithm 1 Aggregated Greedy Equivalence Search

Input: $X^{(n)}$, containing n i.i.d. observations of X
Output: Estimated APDAG A

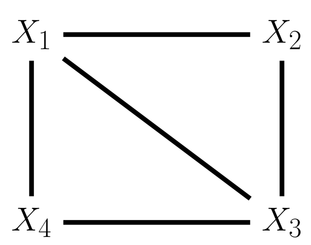
```
 1: Compute the solution path of the sample version of GES for λ ⩾ log(n)/(2n)
 2: Discard all outputs whose skeletons are not contained in the skeleton of the
    output when λ = log(n)/(2n). Denote the remaining CPDAGs, ordered according
    to increasing penalty parameter λ, by Ĉ = {Ĉ_0, ..., Ĉ_k}
 3: A ← Ĉ_0
 4: for i ∈ {1, ..., k} do
 5:   P ← A
 6:   for All edges in Ĉ_i do
 7:     if an edge is oriented in Ĉ_i but not in P then
 8:       Orient it in P as in Ĉ_i
 9:     end if
10:   end for
11:   if P is extendible to a DAG then
12:     A ← P
13:   end if
14: end for
15: A ← MeekOrient(A)
16: return A
```

Example: Concrete example with weights as shown in the weighted DAG and errors $\varepsilon \sim \mathcal{N}(0, D)$ where D is a diagonal matrix with entries $(0.4, 0.3, 0.4, 0.3)$.

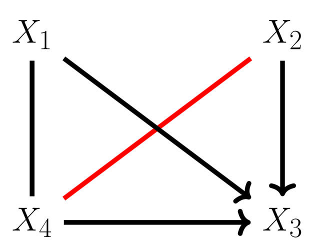
- $\hat{C}_0$ is equal to the true CPDAG. It has no orientations and represent the starting point of the aggregation process.
- We discard two CPDAGs since they contain an edge (in red) not present in $\hat{C}_0$.
- $\hat{C}_1$ and $\hat{C}_2$ contain a v-structure. We add these orientations using $\hat{C}_1$.
- The last two PDAGs show the result before and after applying MeekOrient.



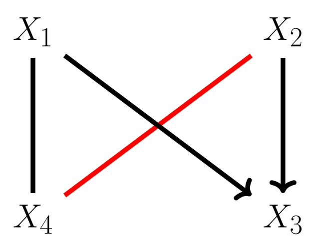
True weighted DAG



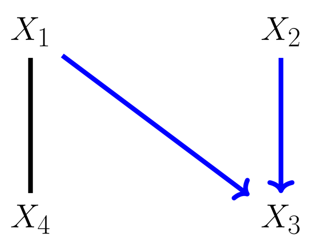
$\hat{C}_0$, no orientations



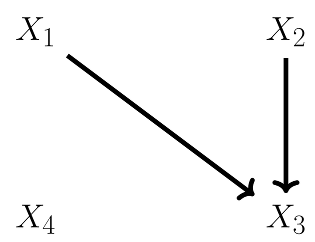
Forbidden edge, discard



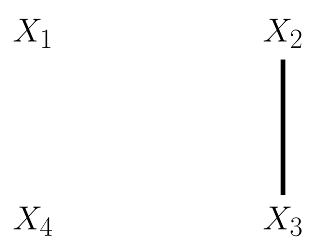
Forbidden edge, discard



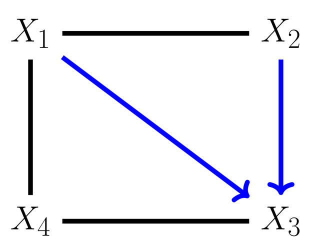
$\hat{C}_1$, new v-structure



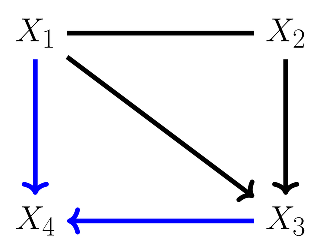
$\hat{C}_2$, same v-structure



$\hat{C}_3$, no new orientations

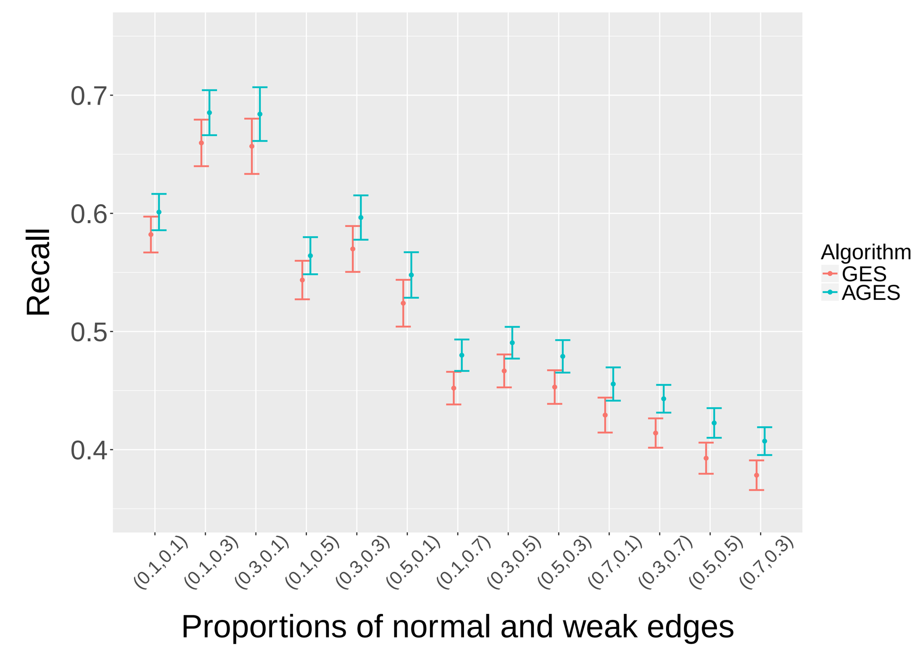


Without MeekOrient



With MeekOrient

Performance: We compare the directed part of the outputs of GES and AGES with respect to the true DAGs.



- The skeleton of GES and AGES are the same by construction.
- We gain in recall in most of the simulation settings (shown here: $n = 10000$, $p = 10$).
- We obtain the same precision as GES in most of the simulation settings (not shown here).
- The monotonicity in λ of the two phases of GES allows for efficient computation. However, the two phases together are not monotone.

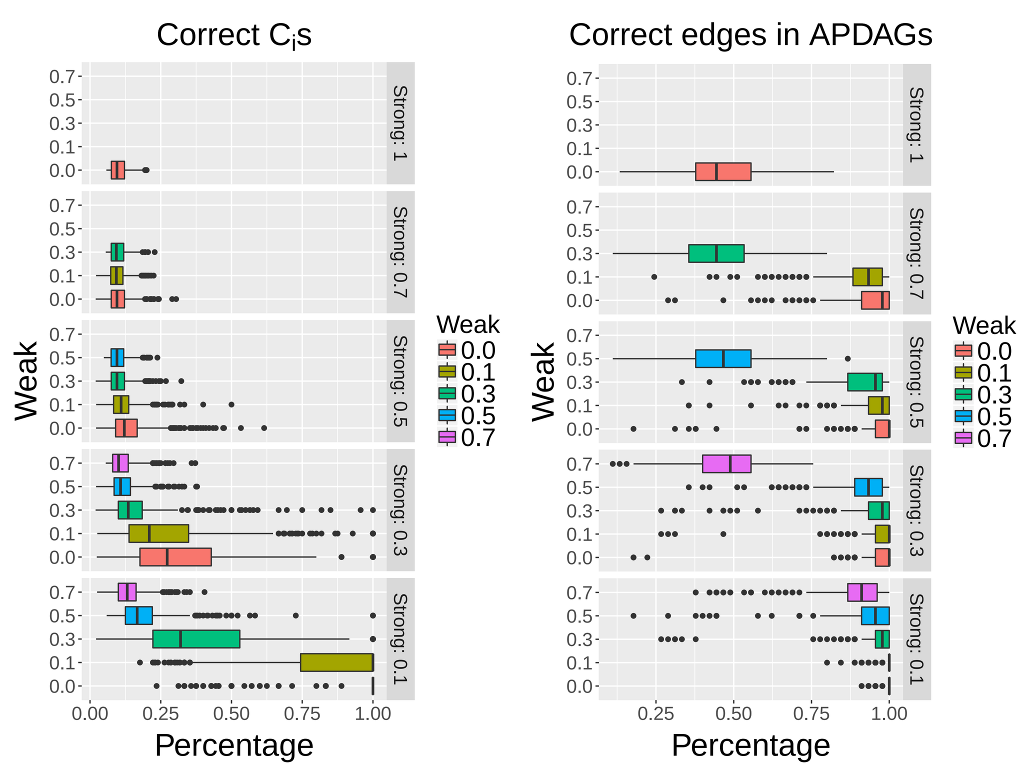
The AGES δ-strong faithfulness condition

Definition (AGES δ-strong faithful) A multivariate Gaussian distribution is said to be *AGES δ-strong faithful* with respect to a DAG $G = (X, E)$ if δ-strong faithfulness holds for every triple (X_i, X_j, S) (X_i, X_j, S as in δ-strong faithfulness) belonging to at least one of the following two sets:

1. Consider the output of the forward phase of oracle GES with penalty parameter $\lambda = -1/2 \log(1 - \delta^2)$. The first set consists of all triples (X_i, X_j, S) that GES considers for a further edge addition.
2. The second set consists of all triples (X_i, X_j, S) that have been used to delete an edge during the backward phase.

Satisfiability of the AGES δ-strong faithfulness condition

- **Left plot: Percentage of correctly found CPDAGs by oracle GES in each solution path (step 3 of the APDAG box).**
- **Right plot: Proportion of correct edge orientations found by oracle AGES as compared to the true APDAGs (step 4 of the APDAG box).**



- AGES δ-strong faithfulness is much easier to satisfy than δ-strong faithfulness, which is basically never satisfied for the shown settings with 10 vertices.
- Correct APDAGs found even more often than suggested from the percentage of correctly found C_is. This happens because not necessarily all CPDAGs in the solution path are used to construct an APDAG.

References

[1] D. M. Chickering. Optimal structure identification with greedy search. *J. Mach. Learn. Res.*, 3:507–554, 2002.

[2] C. Meek. Causal inference and causal explanation with background knowledge. In *Proceedings of UAI 1995*, pages 403–410, 1995.

[3] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. MIT Press, Cambridge, 2nd edition, 2000.