Munich Center for Machine Learning (MCML)
Research group Mathematical Statistics
Technical University of Munich

Munich Center for Machine Learning

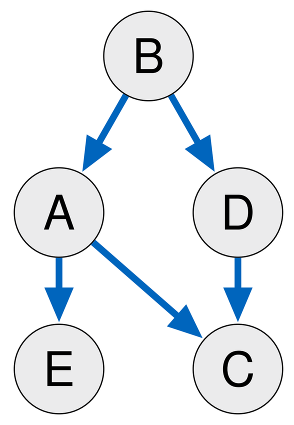
Rank-Based Causal Discovery for Post-Nonlinear Models

Grigor Keropyan [1] , David Strieder[1,2], Mathias Drton[1,2]

1. Motivation

Inferring causal structures among complex systems of interacting variables:

Protein A	Protein B	Protein C	Protein D	Protein E
26.4	3.2	2.5	10.1	162.3
47.3	5.2	3.8	23.4	275.9
31.5	12.6	3	12.2	366.1
55.8	5.2	7.1	26.3	19.6
12.1	4.7	2.1	18.5	136
25.2	41.1	14.2	12.4	105.3
22.1	18.5	2.8	22.6	103.3
19.9	2.8	2.4	17.9	312.6
36.5	4.1	3.8	6.8	25.9
7.1	4.3	8.7	20.9	17.6
6.9	42.2	22.1	11.1	57.8



Example: Inferring protein signaling networks from single-cell data with causal structure learning algorithms.

Structural causal models postulate noisy functional relations among interacting variables.

Post-nonlinear models (PNL) consitute

- a **flexible** subclass of causal models and
- ensure **unique identification** from observational data.

Existing methods learn functional relations by minimizing residual dependencies and subsequently test independence from residuals to determine causal orientations.

⇒ **prone to overfitting** and difficult to tune.

We propose a **new approach that uses rank-based methods** to estimate the functional parameters and therefore disentangles the estimation of the non-linear functions from the independence tests used to find causal orientations.

2. Setup

Flexible causal model that ensures identifiability:

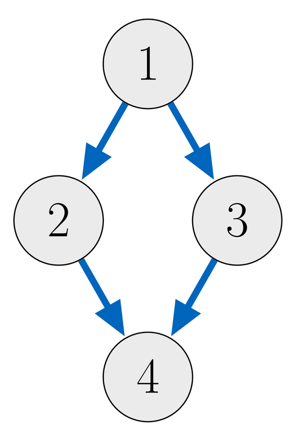
- Observational data of $(X^{(1)}, .., X^{(m)})$ follows **PNL model**, that is, for $k = 1, \ldots, m$

$$X^{(k)} = f^{(k)}\left(g^{(k)}\left(X^{(\mathsf{PA}_k)}\right) + \varepsilon^{(k)}\right).$$

- $f^{(k)}$ is continuous and strictly increasing.
- $\varepsilon^{(k)}$ are jointly independent noise variables.
- Focus on linear $g^{(k)}$. Can be extended to non-linear functions via basis expansions or MLPs.
- **Goal: Infer causal structure**, naturally represented by a **directed acyclic graph**.

Example:

$$X^{(1)} = \varepsilon^{(1)},$$
$$X^{(2)} = f^{(2)}\big(X^{(1)}\beta^{(2)} + \varepsilon^{(2)}\big),$$
$$X^{(3)} = f^{(3)}\big(X^{(1)}\beta^{(3)} + \varepsilon^{(3)}\big),$$
$$X^{(4)} = f^{(4)}\big((X^{(2)}, X^{(3)})\beta^{(4)} + \varepsilon^{(4)}\big),$$



3. Method

Learning causal structures via **recursive sink node identification**:

1. For each node k, **learn functions** $\hat{f}^{(k)}$ and $\hat{\beta}^{(k)}$ corresponding to PNL model
$$\big(f^{(k)}\big)^{-1}(X^{(k)}) = (X^{(-k)})^T\beta^{(k)} + \varepsilon^{(k)}$$
where $X^{(-k)}$ are remaining nodes except k.
2. Use Step 1. to estimate residuals
$$\hat{\varepsilon}^{(k)} = (\hat{f}^{(k)})^{-1}(X^{(k)}) - (X^{(-k)})^T\hat{\beta}^{(k)}.$$
3. **Test for independence** between residuals and remaining nodes $X^{(-k)}$ with consistent test, e.g. HSIC (Gretton et al., *Measuring statistical dependence with hilbertschmidt norms,* 2005).
4. **Identify** node which minimizes the test statistic as **sink node**, remove sink node and repeat from Step 1. to obtain causal ordering.
5. **Prune** redundant edges to obtain structure.

Consistently recovers causal ordering under sink node identifiability and consistency of estimators.

Key Idea: Leverage rank invariances to estimate functional parameters and **disentangle learning** of function **from independence tests** used to find causal orientations.

Post-nonlinear rank regression methods for Step 1. of structure learning procedure:

- Insight: strictly increasing $f^{(k)}$ preserves ranks
- Idea for $\hat{\beta}^{(k)}$: maximize **pairwise rank likelihood** function based on
$$\mathbb{P}\big(X_j^{(k)} > X_i^{(k)} \big| X^{(-k)}\big)$$
$$= \mathbb{P}\big(\varepsilon_j^{(k)} - \varepsilon_i^{(k)} > (X_i^{(-k)} - X_j^{(-k)})^T\beta^{(k)} \big| X^{(-k)}\big)$$
- Idea for $\big(\hat{f}^{(k)}\big)^{-1}(x)$: maximize **smoothed rank correlation** objective function
$$Q(z) = \frac{1}{n(n-1)}\sum_{i \neq j}(d_{jx} - d_{iy}) \times$$
$$\Phi\Big(\sqrt{n}((X_j^{(-k)} - X_i^{(-k)})^T\hat{\beta}^{(k)} - z\Big),$$
with $d_{jx} := \mathbb{1}(X_j^{(k)} \geq x)$ and $d_{iy} := \mathbb{1}(X_i^{(k)} \geq y)$.
- $f^{(k)}$ can be replaced by location-scale transformations. For unique identification, we assume $\big(f^{(k)}\big)^{-1}(y) = 0$ for some known y.

We provide **two methods to recover the causal ordering**, a computationally efficient method RankG under Gaussian noise, and a general method for unknown noise distributions RankS.

4. Simulations

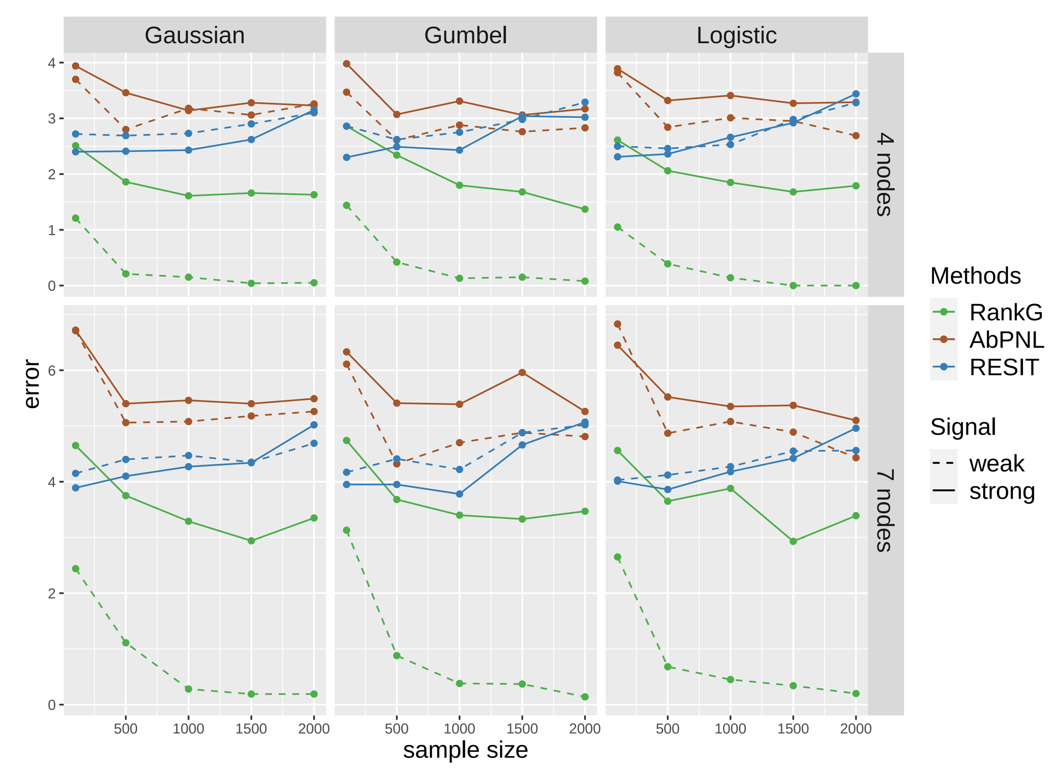
Experiments with synthetic PNL data:

- Randomly selected DAGs on 4 or 7 nodes.
- Different noise distributions.
- Weak and strong signal settings.

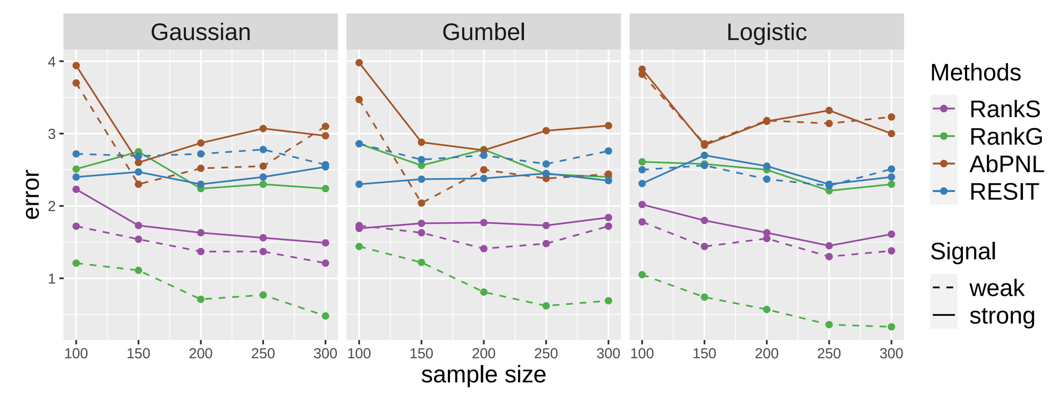
We report **number of wrongly oriented edges** in estimated causal ordering, that is,

$$\#\{(i,j) : \hat{\pi}(i) \to \hat{\pi}(j) \in \mathcal{G} \text{ and } j < i\}.$$

- **Rank-based methods outperform the competition** in all considered settings.
- Consistently recover valid causal ordering in low sample sizes for weak signals.
- High noise compared to signal strength induces more rank changes and thus, better performance in weak signal settings.
- RankG is computationally efficient with computation times comparable with AbPNL method.
- RankS indicates some robustness under noise misspecification.



Mean error against sample size.



Mean error against sample size (4 nodes).

This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 83818), the German Federal Ministry of Education and Research, and the Bavarian State Ministry for Science and the Arts.

[2]Munich Center for Machine Learning (MCML)
[1]Research group Mathematical Statistics
Technical University of Munich, Germany

International Conference on Artificial Intelligence and Statistics
April 25-27, 2023, Valencia, Spain