On the Robustness of Causal Discovery with Additive Noise Models on Discrete Data

Kang Du, Austin Goddard, Yu Xiang

University of Utah, Department of Electrical and Computer Engineering

Motivation

The **Additive Noise Models** (ANMs) framework for causal discovery has gained much attention due to its strong theoretical guarantees, as well as superior empirical performances on a wide range of real-world data. For observational data, however, **quantization (or discretization)** is often an **inevitable** preprocessing step depending on measurement precision requirements. It is thus crucial to understand how **sensitive** the ANMs are with respect to quantization.

Main contributions

- Empirically **examine** the **discrete variants** of the ANM-based causal discovery methods over **synthetic** and **real-world** datasets.
- Propose a **simple** yet **effective** discrete method that is relatively **robust** compared with discrete variants of ANMs.
- Show that the **discrete variants** are **outperformed** by the **original ANM** method developed for continuous data (Hoyer et al., 2009), which is **consistent** with the observations made by Mooij et al., 2016.

Causal discovery procedure

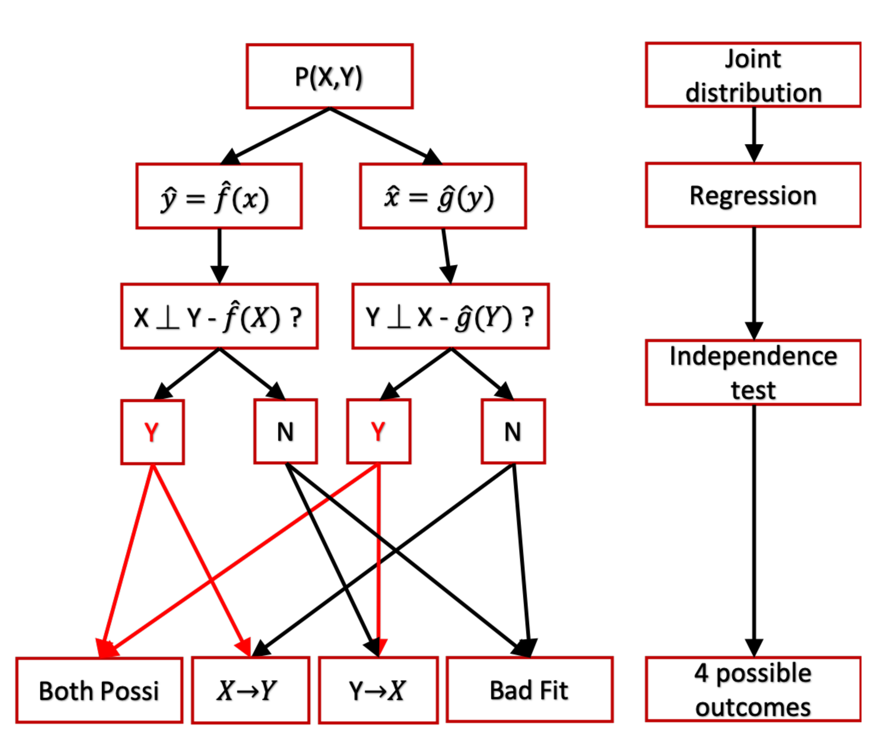

Figure 1: Causal discovery procedure.

Data quantization

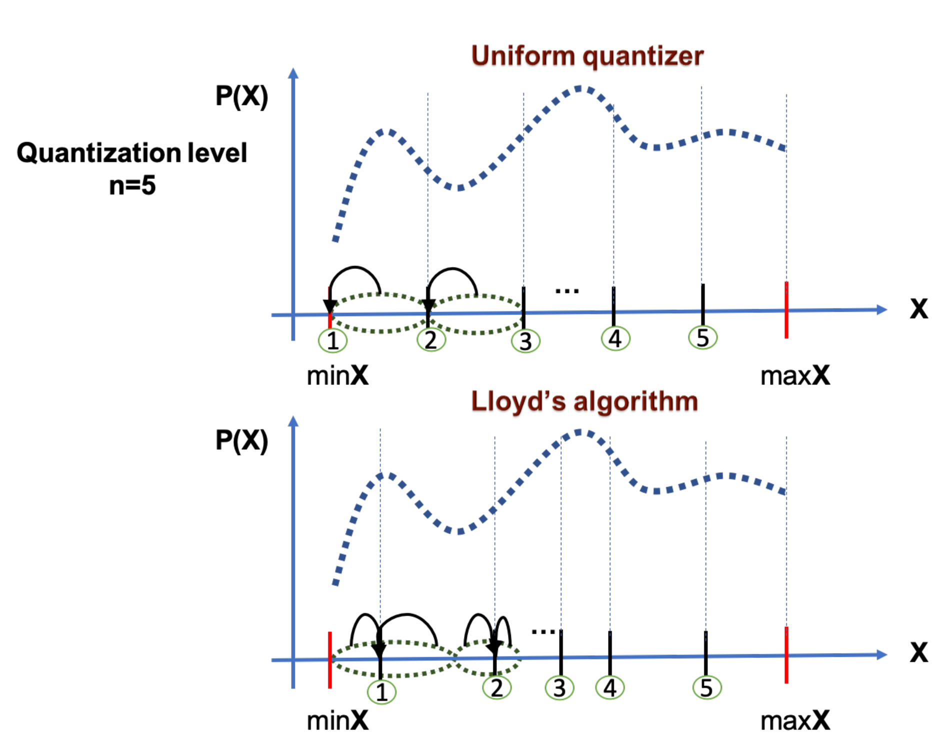

Figure 2: Data quantization.

4+1 Algorithms

4 discrete variants of **existing** ANM-based causal discovery algorithms:

1. *Discrete regression with chi-square test* (DR-chi^2).
2. *Discrete regression with* HSIC *Gaussian kernel* (DR-HSICg).
3. *Discrete regression with* HSIC *discrete kernel* (DR-HSICd).
4. *Gaussian process regression with* HSIC *Gaussian kernel* (GPR-HSICg).

Inspired by the **superior performance** of **algorithm 4**, **we propose**:

5. **Conditional expectation regression** *with* HSIC *Gaussian kernel* (**CER-HSICg**).

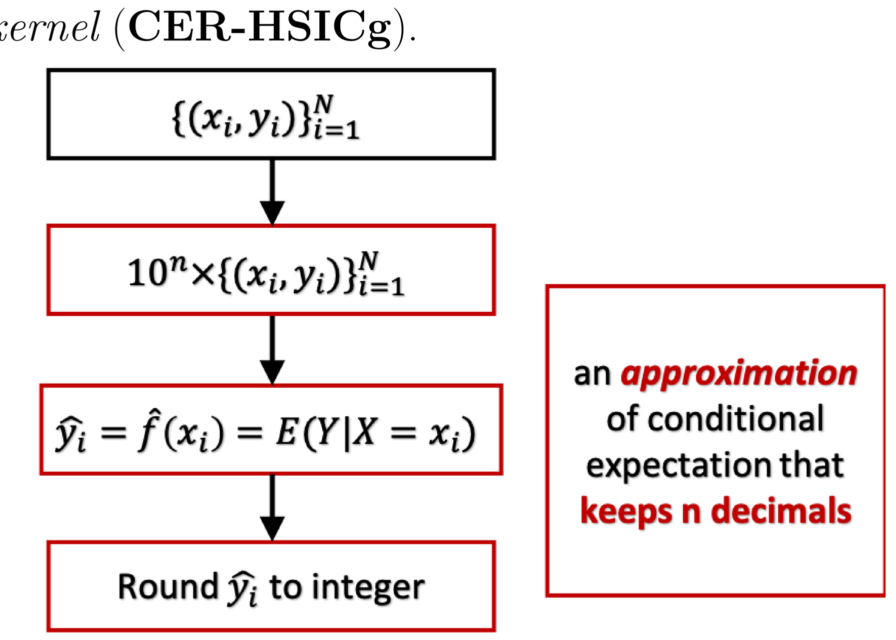

Figure 3: Conditional expectation regression.

The **intuition** for using **conditional expectation regression** as an **alternative method** for **Gaussian process regression** is explained by Figure 4.

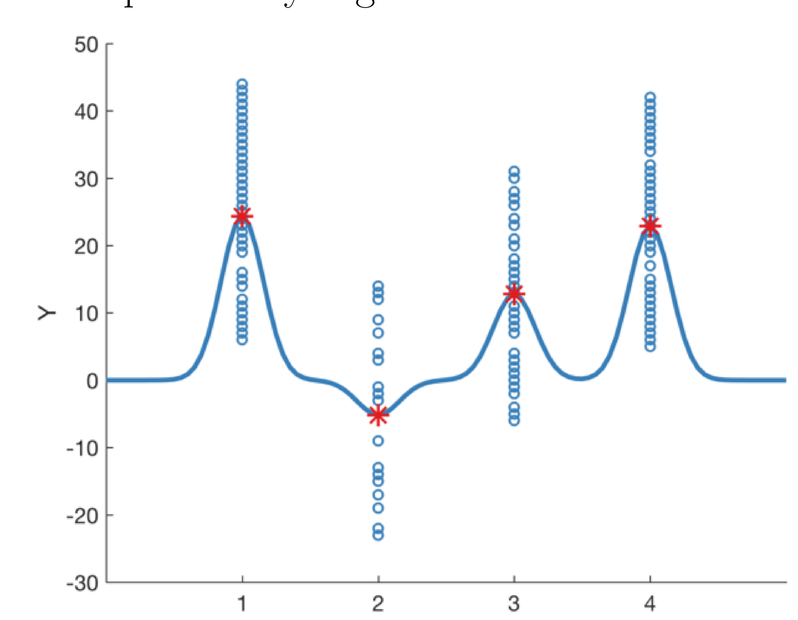

Figure 4: For a sample (of size 1000) from dataset A1 (circles), where each circle represents potentially many repetitions, the output of Gaussian process regression (line) is very **close** to the conditional expectation of Y for each x (star).

Synthetic and real-world Datasets

Dataset (X and Y)	Sample size	suppX	suppY
Synthetic A1	100*1000	≤ 6	≥ 30
Synthetic A2	100*1000	≥ 30	≤ 6
B1: Height and Temperature	727	506	66
B2: Abalone Gender and Length	1000	3	63
B3: Temperature and Hour	1000	33	24
B4: Age and Strength of Concrete	1030	14	845

Figure 5: Dataset information.

Experiments

On synthetic datasets

Table 1: Dataset A1 with uniform discretization ($\alpha = 0.05$).

	Cor. Dir.				Wrong Dir.				Both Dir. Poss.				Bad Fit Both Dir.			
Discr. Level	5	10	15	no	5	10	15	no	5	10	15	no	5	10	15	no
DR-chi^2 (%)	1	1	1	94	0	0	0	0	0	1	1	1	99	98	98	5
DR-HSICd (%)	4	3	3	94	0	0	0	0	0	1	1	1	96	96	96	5
DR-HSICg (%)	25	59	55	52	0	0	1	0	0	1	2	0	75	40	42	48
GPR-HSICg (%)	25	59	55	99	0	0	1	0	0	1	2	0	75	40	42	1
CER-HSICg (%)	13	42	51	82	0	0	0	0	0	1	0	3	87	57	49	15

Table 2: Dataset A1 using Lloyd's algorithm ($\alpha = 0.05$).

	Cor. Dir.				Wrong Dir.				Both Dir. Poss.				Bad Fit Both Dir.			
Discr. Level	5	10	15	no	5	10	15	no	5	10	15	no	5	10	15	no
DR-chi^2 (%)	0	0	0	94	1	1	1	0	0	0	0	1	99	99	99	5
DR-HSICd (%)	0	0	0	94	0	1	1	0	0	0	0	1	100	99	99	5
DR-HSICg (%)	19	56	76	52	0	1	3	0	0	0	3	0	80	43	18	48
GPR-HSICg (%)	14	70	78	99	0	0	0	0	0	0	0	0	86	30	22	1
CER-HSICg (%)	13	42	51	82	0	0	0	0	0	1	0	3	87	57	49	15

On real-world datasets

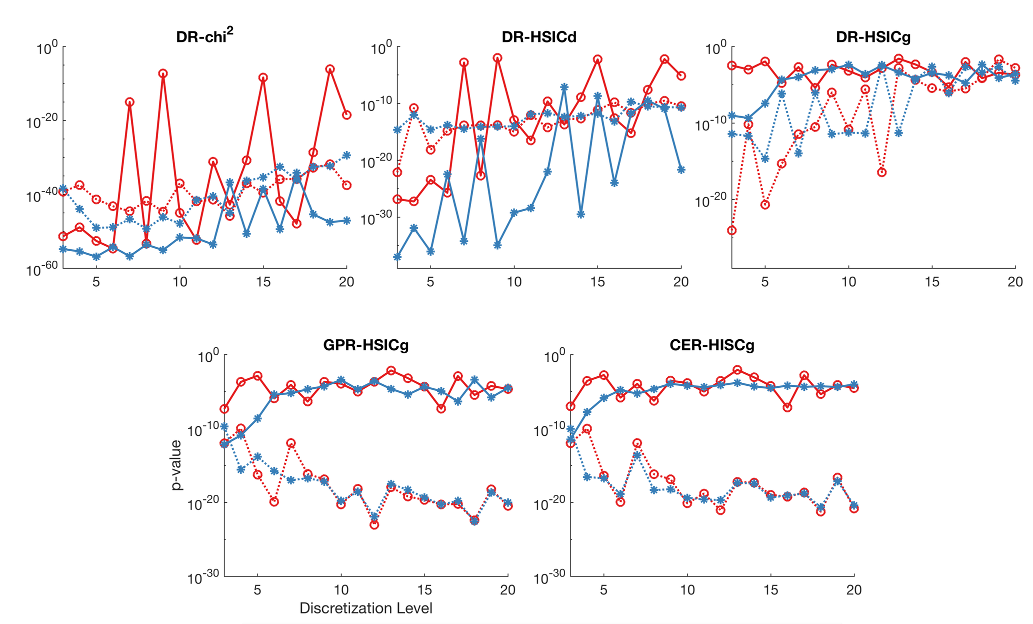

Figure 6: Dataset B1: *height vs. temperature.*

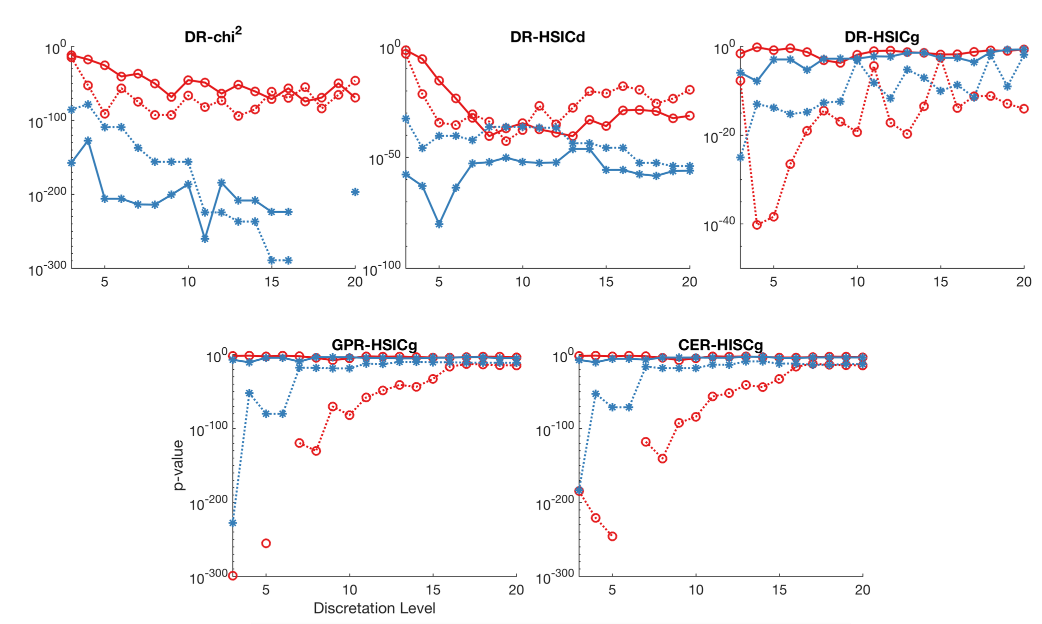

Figure 7: Dataset B2: *gender vs. length.*

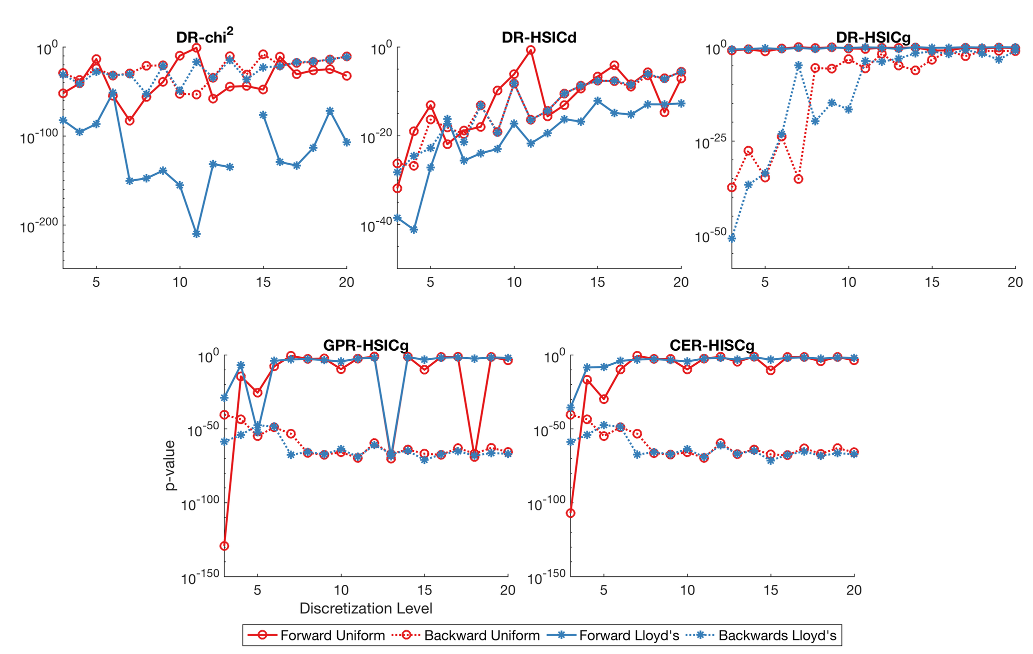

Figure 8: Dataset B3: *hour vs. temperature.*

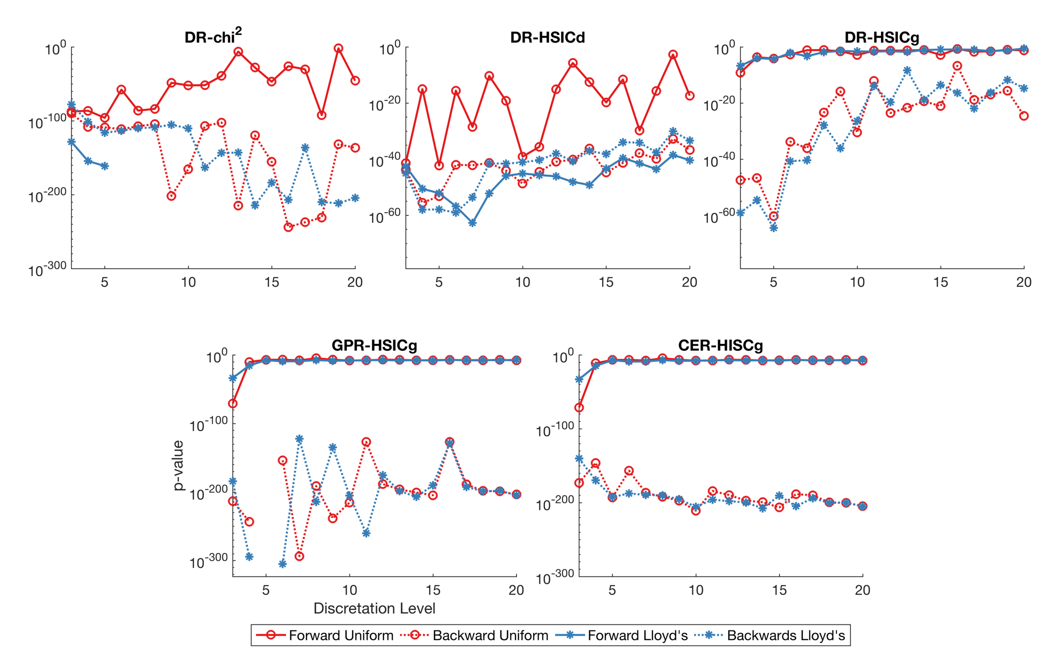

Figure 9: Dataset B4: *concrete age vs. strength.*

Reference

Hoyer, P. O., Janzing, D., Mooij, J. M., Peters, J., & Schölkopf, B. (2009). Nonlinear causal discovery with additive noise models. In *Advances in neural information processing systems.*

Mooij, J. M., Peters, J., Janzing, D., Zscheischler, J., & Schölkopf, B. (2016). Distinguishing cause from effect using observational data: Methods and benchmarks. *The Journal of Machine Learning Research, 17*(1), 1103–1204.

Peters, J., Janzing, D., & Scholkopf, B. (2011). Causal inference on discrete data using additive noise models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33*(12), 2436–2450.