Characteristic Imsets via Quasi-Independence Gluing [1]

Irem Portakal, Joseph Johnson, Ben Hollering, Liam Solus

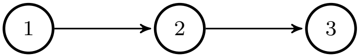
Technical University of Munich

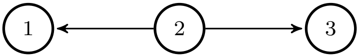
ETH-UCPH-TUM Workshop on Graphical Models
Raitenhaslach
October 13, 2022

[1] arXiv:2209.01834

Causal Discovery and Markov Equivalence

- Directed acyclic graphical (DAG) models are widely used to model conditional independence and causal relationships.
- The causal discovery algorithms typically only recover the Markov equivalence class that best fits the data.
- Verma and Pearl (1992)[2] showed that two DAGs are Markov equivalent if and only if they have the same skeleton and *v-structures (colliders)*.





[2] T. Verma and J. Pearl, An algorithm for deciding if a set of observed independencies has a causal explanation, 1992.

Characteristic Imsets

- Recently, a new geometric perspective on causal discovery algorithms has emerged which uses *characteristic imsets* (CIM) to embed DAGs in Euclidean space.[3]

Definition

Given a DAG $\mathcal{G} = ([n], E)$, the *characteristic imset* is a vector $c_{\mathcal{G}} \in \mathbb{R}^{2^n - n - 1}$ with entries indexed by $S \subseteq [n]$, $|S| \geq 2$ such that

$$(c_{\mathcal{G}})_S := \begin{cases} 1 & \text{if there exists } i \in S \text{ such that } S \setminus \{i\} \subseteq \mathrm{pa}_{\mathcal{G}}(i), \\ 0 & \text{otherwise.} \end{cases}$$

Theorem (Studený-Hemmecke-Lindner 2010)

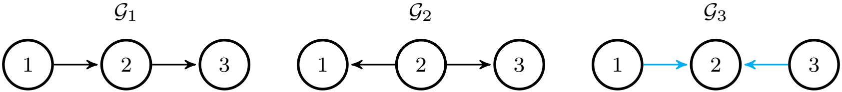
The following are equivalent:

1. $\mathcal{G} = ([n], E)$ *and* $\mathcal{H} = ([n], E')$ *are Markov equivalent,*
2. $c_{\mathcal{G}} = c_{\mathcal{H}}$.

[3] M. Studený, R. Hemmecke, and S. Lindner, Characteristic imset: A simple algebraic representative of a bayesian network structure, 2010.

Example: The path graph P_3

- Consider the DAGs with the underlying path graph P_3:



- For $S \subseteq [3]$ such that $|S| \geq 2$

$$(c_{\mathcal{G}})_S := \begin{cases} 1 & \text{if there exists } i \in S \text{ such that } S \setminus \{i\} \subseteq \mathrm{pa}_{\mathcal{G}}(i), \\ 0 & \text{otherwise.} \end{cases}$$

- Their characteristic imsets are

$$\begin{array}{rcccc} & 12 & 13 & 23 & 123 \\ c_{\mathcal{G}_1} = & (1, & 0, & 1, & 0) \\ c_{\mathcal{G}_2} = & (1, & 0, & 1, & 0) \\ c_{\mathcal{G}_3} = & (1, & 0, & 1, & 1) \end{array}$$

The Characteristic Imset Polytope

- Characteristic imsets allow us to rephrase the problem of causal discovery as a linear programming problems over the polytope

$$\mathrm{CIM}_n := \mathrm{conv}(c_{\mathcal{G}} \in \mathbb{R}^{2^n-n-1} \mid \mathcal{G} = ([n], E) \text{ a DAG})$$

which led to many new geometric perspectives on causal discovery.

- Linusson, Restadh, and Solus[4] showed that the moves used in many classic causal discovery algorithms (e.g. GES, GIES, and MMHC algorithms) to move from one DAG to another correspond to edges of CIM_n.
- They also showed that the polytope

$$\mathrm{CIM}_G := \mathrm{conv}(c_{\mathcal{G}} \mid \mathcal{G} \text{ has skeleton } G)$$

is a face of CIM_n, where G is an undirected graph on n vertices.

[4] S. Linusson, P. Restadh, and L. Solus. Greedy causal discovery is geometric, 2022.

The Characteristic Imset Ideal

- Since CIM_G is a lattice polytope, it has an associated toric ideal I_G which we call the *characteristic imset ideal* of G.
- Let $\mathcal{G}_1, \ldots, \mathcal{G}_n$ be DAGs which represent the Markov equivalence classes of G.
- Then I_G is the kernel of the monomial map

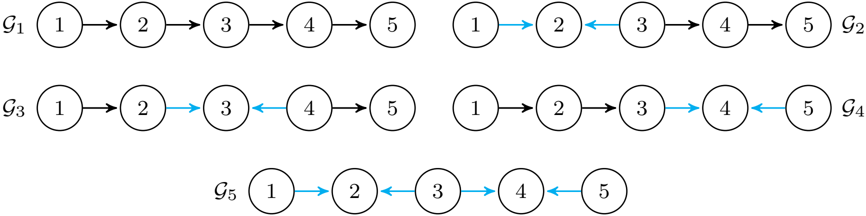
$$\psi_G : \mathbb{R}[z_{\mathcal{G}_1}, \ldots, z_{\mathcal{G}_n}] \to \mathbb{R}[t_S \mid S \subseteq [n], |S| \geq 2]$$
$$z_{\mathcal{G}_i} \mapsto t^{c_{\mathcal{G}_i}}$$

- Understanding the algebraic structure of I_G can help us understand the combinatorial structure of CIM_G.

Problem

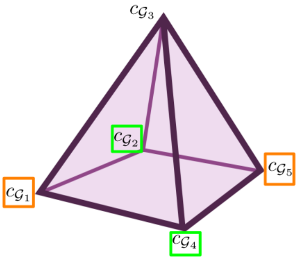
Determine a Gröbner basis for the ideal I_G using the combinatorics of the underlying graph G.

Example: Characteristic imset polytope CIM_{P_5}



$$I_G = \langle z_{\mathcal{G}_1} z_{\mathcal{G}_5} - z_{\mathcal{G}_2} z_{\mathcal{G}_4} \rangle$$

	$\mathcal{G}_1$	$\mathcal{G}_2$	$\mathcal{G}_3$	$\mathcal{G}_4$	$\mathcal{G}_5$
e	1	1	1	1	1
123	0	0	0	1	1
234	0	0	1	0	0
345	0	1	0	0	1

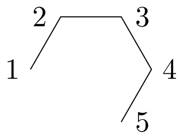


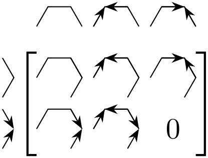
$z_{\mathcal{G}_1} \mapsto t^{c_{\mathcal{G}_1}} = t_e$ $\qquad$ $z_{\mathcal{G}_2} \mapsto t^{c_{\mathcal{G}_2}} = t_e t_{345}$

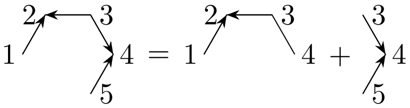
$z_{\mathcal{G}_5} \mapsto t^{c_{\mathcal{G}_5}} = t_e t_{123} t_{345}$ $\qquad$ $z_{\mathcal{G}_4} \mapsto t^{c_{\mathcal{G}_4}} = t_e t_{123}$

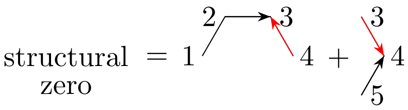
Example

- Consider the path graph P_5. This ideal I_{P_5} in fact comes from a *quasi-independence model*:









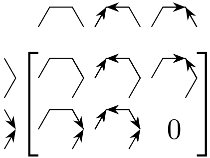
Quasi-Independence Models

- Let X and Y be discrete random variables with states r and s. Quasi-independence models describe the situation in which some combinations of states of X and Y cannot occur together (structural zeros), but X and Y are otherwise independent of one another.
- Let $Q \subseteq [r] \times [s]$ be the subset of states that can occur together.
- The *quasi-independence ideal associated to Q*, denoted I_Q, is the kernel of the monomial map

$$\phi_Q : \mathbb{R}[z_{jk} \mid (j,k) \in Q] \to \mathbb{R}[x_j, y_k \mid j \in [r], k \in [s]]$$

$$z_{jk} \mapsto x_j y_k$$

$$\begin{array}{c} \\ x_\emptyset \\ x_4 \end{array} \begin{array}{c} \begin{array}{ccc} y_\emptyset & y_2 & y_3 \end{array} \\ \begin{bmatrix} z_{\mathcal{G}_1} & z_{\mathcal{G}_2} & z_{\mathcal{G}_4} \\ z_{\mathcal{G}_3} & z_{\mathcal{G}_5} & 0 \end{bmatrix} \end{array}$$



$$I_G = \langle z_{\mathcal{G}_1} z_{\mathcal{G}_5} - z_{\mathcal{G}_2} z_{\mathcal{G}_4} \rangle$$

Quasi-Independence Gluing

- At first glance one might hope that the ideal I_T can be realized as a *toric fiber product* but no grading of the required form can exist.
- To get around this we developed a generalization of the toric fiber product which we call a *quasi-independence gluing*:

Definition

Let $Q \subseteq [r] \times [s]$ and $I \subseteq \mathbb{K}[x]$ and $J \subseteq \mathbb{K}[y]$ be homogeneous ideals. The *quasi-independence gluing* of I and J with respect to the set Q is

$$I \times_Q J := \phi_Q^{-1}(I + J)$$

where $I + J \subset \mathbb{K}[x, y]$ and

$$\phi_Q : \mathbb{K}[z_{jk} \mid (j,k) \in Q] \to \mathbb{K}[x_j, y_k \mid j \in [r], k \in [s]]$$
$$z_{jk} \mapsto x_j y_k$$

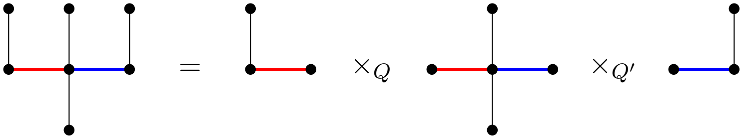
Characteristic Imset Ideals via Quasi-Independence Gluing

Theorem (Hollering - Johnson -  - Solus)

Let $T = ([p], E)$ be a tree, $e = u - v$ be a non-leaf edge of T. Then $I_T = I_{T_u} \times_Q I_{T_v}$ where Q is the set of partings of T.

Corollary (Hollering - Johnson -  - Solus)

Let T be a tree. Then there exists a Gröbner basis of I_T that consists of square-free quadratics (tetrads). Moreover, these quadratics can be explicitly constructed via iterated quasi-independence gluing.



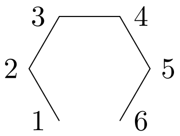
Example: Quasi-independence gluing of P_3 and P_5

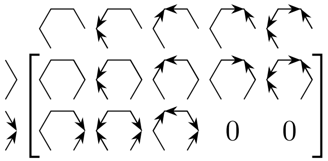
- Let P_3 and P_5 be the paths on 3 and 5 vertices

$$I_{P_3} = \langle 0 \rangle$$
$$I_{P_5} = \langle y_\emptyset y_{24} - y_2 y_4 \rangle$$

- Then $I_{P_6} = I_{P_3} \times_Q I_{P_5} = \phi_Q^{-1}(I_{P_3} + I_{P_5})$
- The quasi-independence ideal associated to this gluing has universal Gröbner basis H_Q generated by the polynomials

$$z_\emptyset z_{25} - z_2 z_5$$
$$z_\emptyset z_{35} - z_3 z_5$$
$$z_2 z_{35} - z_3 z_{25}$$





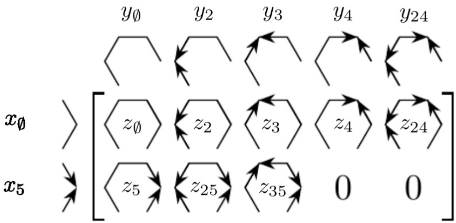
Example: Quasi-independence gluing of P_3 and P_5

- We can find a Gröbner basis for I_{P_6} by *lifting* the only nontrivial generator $y_\emptyset y_{24} - y_2 y_4$ of I_{P_3} and I_{P_5} to

$$z_\emptyset z_{24} - z_2 z_4$$
$$z_5 z_{24} - z_{25} z_4$$

- The characteristic imset ideal (and also its Gröbner basis) of P_6 therefore can be constructed iteratively by quasi-independence gluing of smaller DAGS:

$$\langle z_\emptyset z_{24} - z_2 z_4, z_5 z_{24} - z_{25} z_4, z_\emptyset z_{25} - z_2 z_5, z_\emptyset z_{35} - z_3 z_5, z_2 z_{35} - z_3 z_{25} \rangle$$



Summary

- Understanding the combinatorial structure of characteristic imset polytopes can lead to interesting new results in causal discovery.
- The algebraic structure of the associated toric ideal can help us better understand the combinatorial structure of the polytope.
- *Quasi-independence gluing* is a new operation which generalizes the toric fiber product and can be used to iteratively build larger ideals from smaller ones.
- Characteristic imset ideals of trees are iterated quasi-independence gluings and so we can use this operation to determine a Gröbner basis for I_T.

Question

Are there other families of models that arise from quasi-independence gluing?