UNIVERSITÄT ZU LÜBECK
INSTITUT FÜR
THEORETISCHE INFORMATIK

On Searching for Generalized Instrumental Variables

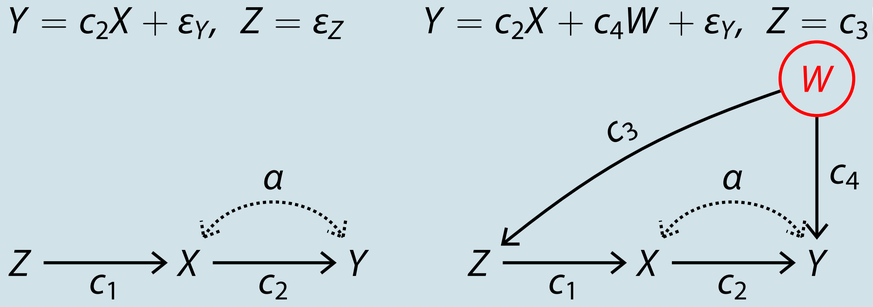
Benito van der Zander and Maciej Liśkiewicz

1. Motivation

A structural equation model (SEM) represents the linear, causal relations between random variables in a graph.

For example $X = c_1 Z + \varepsilon_X$, and $Cov(\varepsilon_X, \varepsilon_Y) = a$,

$Y = c_2 X + \varepsilon_Y$, $Z = \varepsilon_Z$ $\quad$ $Y = c_2 X + c_4 W + \varepsilon_Y$, $Z = c_3 W + \varepsilon_Z$:



(Conditional) instruments are a popular way to calculate the coefficients c_i from observed covariances, e.g.:

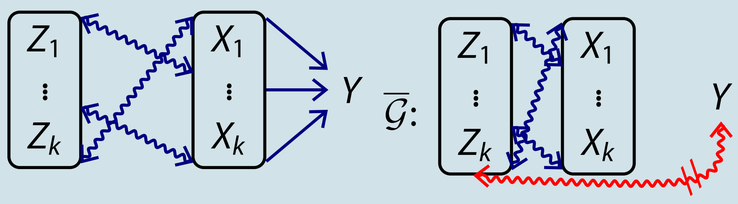
$$c_2 = \frac{Cov(Z,Y)}{Cov(Z,X)} \qquad c_2 = \frac{Cov(Z,Y|W)}{Cov(Z,X|W)}$$

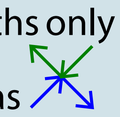
because the first graph implies $Cov(Z,Y) = c_1 c_2$ and $Cov(Z,X) = c_1$.

Brito[1, 2, 3] introduces Instrumental Sets that can allow identification of multiple parameters $X_1 \overset{c_1}{\to} Y, \ldots, X_k \overset{c_k}{\to} Y$ between a node Y and some of its parents $\mathbf{X}$ simultaneously, even if none of these parameters can be identified directly by (conditional) instruments:

Simple Instrumental Set

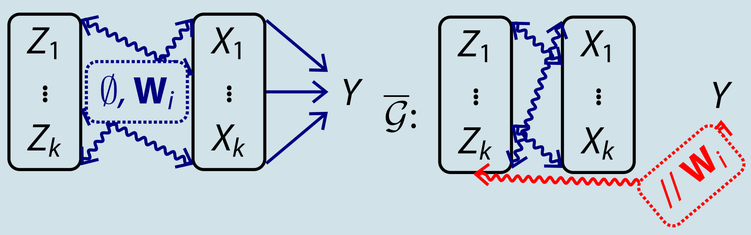
Set $\mathbf{Z}$ is a simple instrumental set for Y and a subset $\mathbf{X}$ of parents of Y, if $\mathbf{X}$ and $\mathbf{Z}$ are d-connected, and $\mathbf{X}$ and Y are not d-connected in $\overline{\mathcal{G}}$:

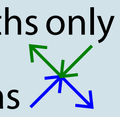


and all paths only intersect as 

Generalized instrumental set

Set $\mathbf{Z}$ is a generalized instrumental set for Y and a subset $\mathbf{X}$ of parents of Y, if $\mathbf{X}$ and $\mathbf{Z}$ are d-connected, and $\mathbf{X}$ and Y can be d-separated by non-descendant sets $\mathbf{W}_1, \ldots, \mathbf{W}_k$ that do not block the paths $\mathbf{X} \sim \mathbf{Z}$:

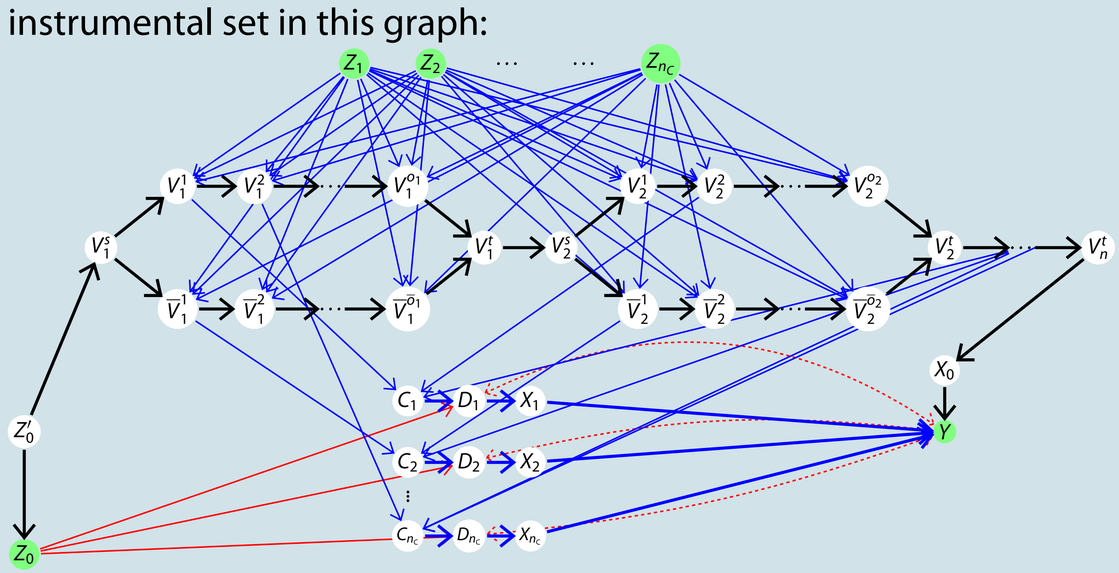


and all paths only intersect as 

So a Simple Instrumental Set is a Generalized Instrumental Set with the constraint $\mathbf{W}_1 = \ldots = \mathbf{W}_k = \emptyset$.

2. NP-Completeness

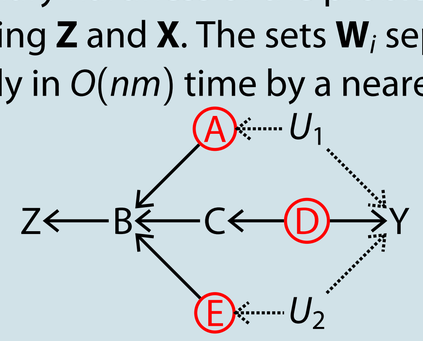
A given 3-SAT instance is solvable, if and only if $\mathbf{Z}$ is an generalized instrumental set in this graph:



where we have a single horizontal path $Z_0 \leftarrow Z'_0 \rightarrow V_1^s \rightarrow \ldots V_k^t \rightarrow Y$ and a vertical path $Z_i \rightarrow V_{v_i}^{u_i} \rightarrow C_i \rightarrow D_i \rightarrow X_i$ for every clause i satisfied by literal $V_{v_i}^{u_i}$

3. Algorithm: Nearest separator

The primary hardness of the problem stems from finding the paths connecting $\mathbf{Z}$ and $\mathbf{X}$. The sets $\mathbf{W}_i$ separating Z_i and Y are found efficiently in $O(nm)$ time by a nearest separator [4].

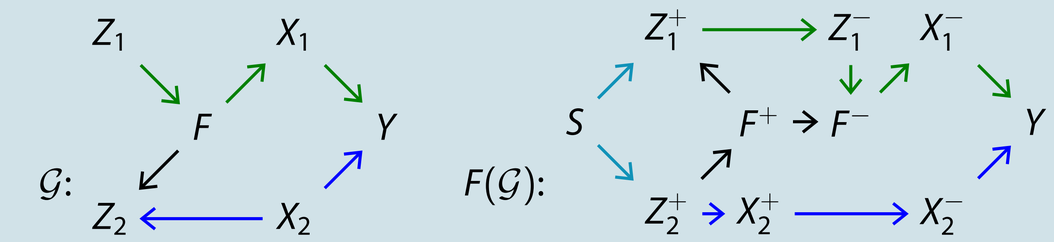


Repeatedly blocking paths between Y and Z with the node closest to Y

It is advantageous to represent bidirectional edges V  W as $V \leftarrow U \rightarrow W$ with an unobservable variable U.

4. Algorithm: Constrained structure

Testing an Instrumental Set with $\mathbf{W}_1 = \ldots = \mathbf{W}_k = \mathbf{W}$ is possible in $O(nm)$ time by transforming the causal graph to a two-layered flow graph, which converts d-paths to directed paths:

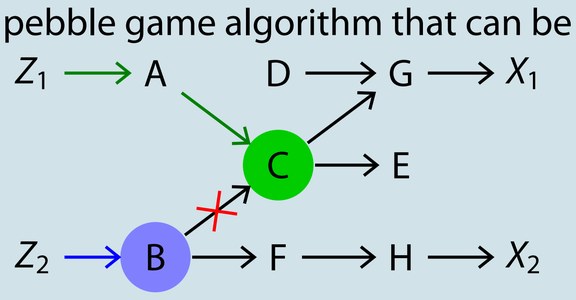


$\mathbf{W}$ is found by the nearest separator algorithm and removed from the flow graph. The non-connectedness of $\mathbf{Z}$ and Y can be tested by a Bayes-Ball algorithm.

A Simple Instrumental Set can also be *found* by this algorithm in $O(nm)$ by constructing the flow graph for a candidate superset $\mathbf{Z} = De(An(\mathbf{X})) \setminus De(An(Y))$, i.e. all nodes reachable from $\mathbf{X}$ and not reachable from Y.

5. Algorithm: Constrained size, constant k

After fixing the $\mathbf{W}_i$ to nearest separators the problem is reduced to finding k d-paths between $\mathbf{Z}$ and $\mathbf{X}$ that only intersect as  . A d-path has either the form $Z \rightarrow X$, $Z \leftarrow X$ or $Z \leftarrow F \rightarrow X$, so guessing the forks F (there are only $O(n^k)$ which is polynomial for fixed k), reduces the problem to finding directed paths intersecting in a certain way. Finding disjoint paths in a DAG is a known problem solvable by a pebble game algorithm that can be extended to our case:



Pebbles move from their starts to their targets.

Highest pebble moves first and can only intersect certain other pebbles.

There are only polynomial many ($O(n^k)$) pebble arrangements, so the search graph of all game moves has polynomial size and tells us if a solution exists. The traces of the pebbles in the original graph are the searched paths.

6. Conclusion and future work

Identification based on Generalized Instrumental Sets is an NP-complete problem that can nevertheless be solved in polynomial time for certain cases. The statistical properties of estimators using those Instrumental Sets should be investigated.

7. References

[1] C. Brito and J. Pearl. In Proceedings of UAI. 2002.
[2] C. Brito. *In Heuristics, Probability and Causality. A Tribute to Judea Pearl*. 2010.
[3] C: Brito. *Graphical Methods for Identification in Structural Equation Models*. PhD thesis, 2004.
[4] B. van der Zander, J. Textor, and M. Liskiewicz. In Proceedings of IJCAI. 2015.