MASTER THESIS PROPOSAL

LEARNING CAUSAL GRAPHICAL MODELS

JOSE M. PEÑA
STIMA, IDA, LIU
JOSE.M.PENA@LIU.SE

1. INTRODUCTION

An alternative acyclic directed mixed graph (aADMG) is a graph with possibly directed and undirected edges but without directed cycles, i.e. $A \rightarrow \cdots \rightarrow A$ is forbidden. There may be up to two edges between any pair of nodes, but in that case the edges must be different and one of them must be undirected to avoid directed cycles. Edges between a node and itself are not allowed. See Figure 1 (left) for an example. As argued by Peña [2016], aADMGs are suitable for representing causal models with additive noise, i.e. each node A is such that $A = g(Pa(A)) + U_A$ where $Pa(A)$ are the parents of A in the aADMG (i.e. the observed causes of A) and U_A is an error term (i.e. the unobserved causes of A). The error nodes can be made explicit by magnifying the aADMG as shown in Figure 1 (right).

Peña [2017a] presents an algorithm for causal effect identification from aADMGs, i.e. an algorithm to compute causal effects from observed quantities. The algorithm assumes that the aADMG at hand is correct. Therefore, developing algorithms for learning aADMGs from observational data is of paramount importance. Peña [2017b] presents a learning algorithm that builds on the ideas by Hoyer et al. [2009], i.e. it exploits the nonlinearities in the data to identify the directions of the causal relationships. Specifically, consider two random variables X and Y that are causally related as $Y = g(X) + U_Y$. Assume that there is no confounding, selection bias or feedback loop, which implies that X and U_Y are independent. Hoyer et al. prove that if the function g is nonlinear, then the correct direction of the causal relationship between X and Y is generally identifiable from observational data: X and U_Y are independent for the correct direction, whereas Y and U_X are dependent for the incorrect direction. This leads to the following causal discovery algorithm. If X and Y are independent then they are not causally related because we assumed no confounding, selection bias or feedback loop. If they are dependent then first construct a nonlinear regression of Y on X to get an estimate $\hat{g}$ of g, then compute the error $\hat{U}_Y = Y - \hat{g}(X)$, and finally test whether X and $\hat{U}_Y$ are independent. If they are so then accept the model $X \rightarrow Y$. Repeat the procedure for the model $Y \rightarrow X$. When both models or no model is accepted, it may be indicative that the assumptions do not hold. Hoyer et al. also propose a generalization to more than two variables: Given a directed acyclic graph (DAG) over the observed random variables, first construct a nonlinear regression of each node on its parents, then compute each node's error, and finally test whether these

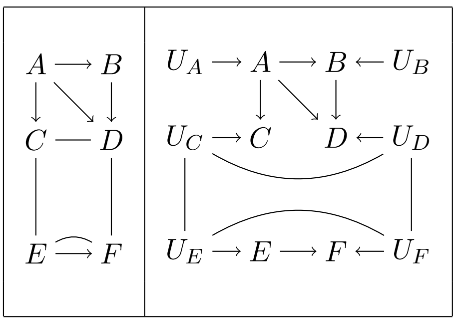

FIGURE 1. Example of an aADMG (left) and its magnification (right).

errors are mutually independent. If they are so then accept the DAG, otherwise reject it. The algorithm performs well in practice [Peters et al., 2014]. Peña [2017b] further generalizes this idea by dropping the assumption that the errors are independent. This can be done by simply using aADMGs instead of DAGs to represent causal models. The parents of a node in an aADMG are its observed causes, whereas its unobserved causes are summarized by an error node that is represented implicitly in the aADMG. We can interpret the undirected edges in the aADMG as the correlation relationships between the different error nodes. The causal structure is constrained to be a DAG, whereas the correlation structure can be any UG. This may be best understood by making the error nodes explicit by magnifying the aADMG (recall Figure 1). The algorithm proposed by Peña basically consists of the following steps: Given an aADMG, first construct a nonlinear regression of each node on its parents, then compute each node's error, and finally test whether these errors satisfy the independences represented by the undirected subgraph of the aADMG. The algorithm has shown promising results for learning chain graphs, i.e. a subclass of aADMGs.

2. Objective

For your master thesis, we propose that you extend the models and/or learning algorithm in Peña [2017b]. In order to get acquainted with the topic, you may want to start by evaluating the learning algorithm developed by Peña [2017b] but this time for unconstrained aADMGs. You may need to modify slightly the implementation by Peña, which is publicly available. This evaluation exercise should give you insights into the weaknesses of the models and/or learning algorithm, which you can address in your thesis.

Peña also presents a second learning algorithm which assumes that the noise is not only additive but normally distributed. If time permits, we also propose that you repeat the exercise above for this second algorithm.

For the evaluation of the algorithms, you can use the same dataset as Peña, namely the DWD dataset which contains climate data from the German Weather Service. The data consists of 349 instances, each corresponding to a weather station in Germany.

References

P. O. Hoyer, D. Janzing, J. Mooij, J. Peters, and B. Schölkopf. Nonlinear Causal Discovery with Additive Noise Models. In *Advances in Neural Information Processing Systems 21*, pages 689–696, 2009.

J. M. Peña. Alternative Markov and Causal Properties for Acyclic Directed Mixed Graphs. In *Proceedings of the 32nd Conference on Uncertainty in Artificial Intelligence*, pages 577–586, 2016.

J. M. Peña. Causal Effect Identification in Alternative Acyclic Directed Mixed Graphs. In *Proceedings of Machine Learning Research*, volume 73, pages 21–32, 2017a.

J. M. Peña. Learning Causal AMP Chain Graphs. In *Proceedings of Machine Learning Research*, volume 73, pages 33–44, 2017b.

J. Peters, J. M. Mooij, D. Janzing, and B. Schölkopf. Causal Discovery with Continuous Additive Noise Models. *Journal of Machine Learning Research*, 15:2009–2053, 2014.