Learning Multivariate Regression Chain Graphs under Faithfulness

Dag Sonntag
ADIT, IDA, Linköping University, Sweden
dag.sonntag@liu.se

Jose M. Peña
ADIT, IDA, Linköping University, Sweden
jose.m.pena@liu.se

Abstract

This paper deals with multivariate regression chain graphs, which were introduced by Cox and Wermuth (1993, 1996) to represent linear causal models with correlated errors. Specifically, we present a constraint based algorithm for learning a chain graph a given probability distribution is faithful to. We also show that for each Markov equivalence class of multivariate regression chain graphs there exists a set of chain graphs with a unique minimal set of lines. Finally, we show that this set of lines can be identified from any member of the class by repeatedly splitting its connectivity components according to certain conditions.

Keywords: Chain Graph, Multivariate Regression Chain Graph, Learning, Bidirected Graph

1 Introduction

In this paper we deal with multivariate regression chain graphs (CGs) which were introduced by Cox and Wermuth (1993, 1996). Graphically Cox and Wermuth represent these CGs with dashed edges to distinguish them from other interpretations, e.g. LWF (Lauritzen, 1996) or AMP (Andersson et al., 2001). Multivariate regression CGs also coincide with the acyclic directed mixed graphs without semi-directed cycles presented by Richardson (2003). A fourth interpretation of CGs can also be found in Drton (2009). The different interpretations of CGs have different merits, but none of the interpretations subsumes another interpretation (Drton, 2009).

The multivariate regression CG interpretation still misses some fundamental elements found and proven in the LWF and AMP interpretations. In this article we study and describe two of these elements. The first is a learning algorithm for learning a CG from a probability distribution faithful to a multivariate regression CG. Similar algorithms have been presented for the LWF (Studený, 1997; Ma et al., 2008) as well as the AMP (Peña, 2012) interpretations. The algorithm presented is constraint based and resembles both Studený's and Peña's algorithms. The second element we present is a feasible split operation and a feasible merge operation similar to the ones presented for the LWF and AMP interpretations (Studený et al., 2009). These splits and mergings can be used to alter the structure of a multivariate regression CG in such a way that it does not change the CG's Markov equivalence class. Finally we show that for each Markov equivalence class of multivariate regression CGs there exists a set of CGs with a unique minimal set of lines. We also show that this set of CGs can be reached by applying feasible splits to any CG in the same Markov equivalence class.

The rest of the paper is organised as follows. Section 2 reviews the concepts used in the rest of the article. Section 3 presents the feasible split and merge operations. Section 4 presents the learning algorithm and proves its correctness.

Section 5 closes the article with some discussion.

2 Preliminaries

In this section, we review some concepts from probabilistic graphical models that are used later in this paper. All the graphs and probability distributions in this paper are defined over a finite set of variables V. With $|V|$ we mean the number of variables in the V and with $|V_G|$ the number of variables in the graph G. Throughout the paper the intended meaning of CGs is multivariate regression CGs if no other interpretation is mentioned. To allow more readable figures bidirected edges are used instead of dashed edges or lines. To not confuse the reader these edges will also be denoted bidirected edges throughout the article.

If a graph G contains an edge between two nodes V_1 and V_2, we write that $V_1 \to V_2$ is in G for a directed edge, $V_1 \leftrightarrow V_2$ is in G for a bidirected edge, and $V_1 - V_2$ is in G for an undirected edge. With $V_1 \multimap V_2$ we mean that either $V_1 \to V_2$ or $V_1 \leftrightarrow V_2$ is in G. With $V_1 \multimap V_2$ we mean that either $V_1 \to V_2$ or $V_1 - V_2$ is in G. With $V_1 \circ\!\!-\!\!\circ V_2$ we mean that there exists an edge between V_1 and V_2 in G. A set of nodes is said to be complete if there exists edges between all pairs of nodes in the set. A complete set of nodes is said to be a clique if there exists no superset of it that is complete.

The parents of a set of nodes X of G is the set $pa_G(X) = \{V_1 | V_1 \to V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The children of X is the set $ch_G(X) = \{V_1 | V_2 \to V_1$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The spouses of X is the set $sp_G(X) = \{V_1 | V_1 \leftrightarrow V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The adjacents of X is the set $adj_G(X) = \{V_1 | V_1 \to V_2, V_1 \leftarrow V_2$, $V_1 \leftrightarrow V_2$ or $V_1 - V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. A path from a node V_1 to a node V_n in G is a sequence of distinct nodes $V_1, \dots, V_n$ such that $V_i \in adj_G(V_{i+1})$ for all $1 \le i < n$. The length of a path is the number of edges in the path. A path is called a cycle if $V_n = V_1$. A path is called descending if $V_i \in pa_G(V_{i+1}) \cup sp_G(V_{i+1})$ for all $1 \le i < n$. The descendants of a set of nodes X of G is the set $de_G(X) = \{V_n |$ there is a descending path from V_1 to V_n in G, $V_1 \in X$ and $V_n \notin X\}$. A path is called strictly descending if $V_i \in pa_G(V_{i+1})$ for all $1 \le i < n$. The strict descendants of a set of nodes X of G is the set $sde_G(X) = \{V_n |$ there is a strict descending path from V_1 to V_n in G, $V_1 \in X$ and $V_n \notin X\}$. The ancestors (resp. strict ancestors) of X is the set $an_G(X) = \{V_1 | V_n \in de_G(V_1), V_1 \notin X, V_n \in X\}$ (resp. $san_G(X) = \{V_1 | V_n \in sde_G(V_1), V_1 \notin X, V_n \in X\}$). Note that the definition for strict descendants given here coincides to the definition of descendants given by Richardson (2003). Our definition of descendants is however needed for certain proofs.

A cycle is called a semi-directed cycle if it is descending and $V_i \to V_{i+1}$ is in G for some $1 \le i < n$. A CG is a graph containing only directed and bidirected edges with no semi-directed cycles. An undirected graph is said to be chordal if every cycle of length four or more has an edge between two non-consecutive vertices in the cycle. A CG is said to be connected if there exists a path between every pair of nodes in it. A connectivity component C of a CG is a maximal set of nodes (wrt to inclusion) st there exists a path between every pair of nodes in C containing only bidirected edges. The connectivity component of a node X in a CG G, denoted $co_G(X)$, is the connectivity component in G to which X belongs. A subgraph of G is a subset of nodes and edges in G. A subgraph of G induced by a set of its nodes X is the graph over X that has all and only the edges in G whose both ends are in X.

A node C is a collider between two nodes A and B in a CG G if there exists edges $A \circ\!\!\to C \leftarrow\!\!\circ B$ in G. An unshielded collider is a collider where $A \notin adj_G(B)$ and we then say that A and B have an unshielded collider over C. With a non-collider node C between two nodes A and B we mean that $A \circ\!\!-\!\!\circ C \multimap B$ is in G.

Let X, Y and Z denote three disjoint subsets of nodes in a CG G. X is said to be separated from Y given Z iff there exists no path between any node in X and any node in Y st: (1) every non-collider on the path is not in Z and (2) every collider on the path is in Z or in $san_G(Z)$. We denote this by $X \perp_G Y | Z$. Likewise, we denote by $X \perp_p Y | Z$ that X is independent of Y

given Z in a probability distribution p. The independence model induced by G, denoted as $I(G)$, is the set of separation statements $X \perp_G Y|Z$.

We say that a probability distribution p is Markovian with respect to a CG G when $X \perp_p Y|Z$ if $X \perp_G Y|Z$ for all X, Y and Z disjoint subsets of V. We say that p is faithful to G when $X \perp_p Y|Z$ iff $X \perp_G Y|Z$ for all X, Y and Z disjoint subsets of V. We say that two CGs G and H are Markovian equivalent or that they are in the same Markov equivalence class iff $I(G) = I(H)$. If G and H have the same adjacencies and unshielded colliders, then $I(G) = I(H)$ (Wermuth and Sadeghi, 2012, Theorem 1).

3 Feasible split and merging

In this section we present the feasible split and feasible merge operations. We also show that there exists a set of CGs with a unique minimal set of bidirected edges for each Markov equivalence class and that these CGs can be reached by repeatedly applying feasible splits to any CG in that Markov equivalence class.

Definition 1. Feasible Split

Let C denote a connectivity component of G and U and L two disjoint subsets of C st $C = U \cup L$ and the subgraph of G induced by U is connected. A split of C, performed by replacing every edge $X \leftrightarrow Y$ with $X \rightarrow Y$ st $X \in U$ and $Y \in L$, is feasible iff:

1 For all $A \in sp_G(U) \cap L$, $U \subseteq sp_G(A)$ holds
2 For all $A \in sp_G(U) \cap L$, $pa_G(U) \subseteq pa_G(A)$ holds
3 For all $B \in sp_G(L) \cap U$, $sp_G(B) \cap L$ is a complete set

---

Definition 2. Feasible Merging

Let U and L denote two connectivity components of G. A merge between the two components, performed by replacing every edge $X \rightarrow Y$ with $X \leftrightarrow Y$ st $X \in U$ and $Y \in L$, is feasible iff:

1 For all $A \in ch_G(U) \cap L$, $pa_G(U) \cup U \subseteq pa_G(A)$ holds
2 For all $B \in pa_G(L) \cap U$, $ch_G(B) \cap L$ is a complete set
3 $de_G(U) \cap pa_G(L) = \emptyset$

Lemma 1. *A CG is in the same Markov equivalence class before and after a feasible split.*

Proof. Let G be a CG and G' a graph st G' is G with a feasible split performed upon it. G and G' are in different Markov equivalence classes or G' is not a CG iff (1) G and G' do not have the same adjacencies, (2) G and G' do not have the same unshielded colliders or (3) G' contains a semi-directed cycle.

First it is clear that the adjacencies are the same before and after the split since the split does not change the adjacencies in G. Secondly let us assume that G and G' do not have the same unshielded colliders. It is clear that a split does not introduce any new unshielded collider, which means that an unshielded collider is removed during the split. Let us say that this unshielded collider is between two nodes X and Y over Z in G st X and/or Y are in L and $Z \in U$. Without loss of generality, let us say that X is in L. $X \leftrightarrow Z$ and $Y \multimap\!\!\!\rightarrow Z$ must then hold in G but $X \notin adj_G(Y)$. If $Y \rightarrow Z$ is in G this does not fulfill constraint 2 in definition 1, hence $Y \leftrightarrow Z$ must hold in G. Now Y can be either in U or L. If $Y \in U$ constraint 1 in definition 1 is violated and if $Y \in L$ constraint 3 in definition 1 is violated. Hence we have a contradiction. Finally let us assume a semi-directed cycle is introduced. This can happen iff we have two nodes X and Y st $X \in de_{G'}(Y)$, $X \in U$ and $Y \in L$. We know no semi-directed cycle existed in G before the split. Then $de_{G'}(Y) \subseteq de_G(Y) \setminus U$ and by definition $X \notin de_G(Y) \setminus U$. Hence we have a contradiction. □

Lemma 2. *A CG is in the same Markov equivalence class before and after a feasible merging.*

Proof. Let G be a CG and G' a graph st G' is G with a feasible merging performed upon it. G and G' are in different Markov equivalence classes or G' is not a CG iff: (1) G and G' do not have the same adjacencies, (2) G and G' do not have the same unshielded colliders or (3) G' contains a semi-directed cycle.

First it is clear that the adjacencies are the same before and after the merging since the merging does not change the adjacencies in G. Secondly let us assume that G and G' do not have the same unshielded colliders. It is clear that a merging does not remove any unshielded colliders which means that there has to exist an unshielded collider between two nodes X

and Y over Z in G' that does not exist in G. It is clear that the following must hold: $X \notin adj_G(Y)$, $Z \in U$, X and/or $Y \in L$. Without loss of generality, let us say that $X \in L$, which gives us that $X \in ch_G(Z)$. If $Y \in L$ we have $Y \in ch_G(Z)$ which contradicts constraint 2 in definition 2. If $Y \notin L$ we have that $Y \in pa_G(Z)$ or $Y \in sp_G(Z)$, either contradicts constraint 1 in definition 2. Hence an unshielded collider can not be removed. Finally let us assume a semi-directed cycle is introduced. This means that we have three nodes X, Y and Z st $X \in L, Z \in U, Y \notin U \cup L$, $Z \leftrightarrow X \leftarrow Y$ is in G' and $Y \in de_{G'}(Z)$. However, this violates condition 3 in definition 2, because $Y \in pa_G(X)$ and $Y \in de_G(U)$. □

We will now show that there exists a set of CGs which have a unique minimal set of bidirected edges for each Markov equivalence class. We also show that this set of edges is shared by all CGs in the class and that the CGs containing no other bidirected edges than the minimal set, can be reached by repeatedly performing feasible splits on any CG in the class.

Theorem 1. *For a Markov equivalence class of CGs, there exists a unique minimal (wrt inclusion) set of bidirected edges that is shared by all members of the class.*

Proof. Assume to the contrary that there exists two CGs G and G' st $I(G) = I(G')$ and G and G' have two different minimal sets of bidirected edges. Now for all ordered pair of nodes A and B st $A \leftrightarrow B$ is in G but $A \rightarrow B$ is in G', replace $A \leftrightarrow B$ in G with $A \rightarrow B$ and call this new CG H. Obviously H has a proper subset of the bidirected edges in G. We can also see that H has no semi-directed cycle, because if it had a semi-directed cycle this cycle would also have been in G or G' which contradicts that G and G' are CGs. Finally we can see that $I(H) = I(G)$ because H has the same adjacencies and unshielded colliders as G. To see the last, note that it is impossible that an unshielded collider $C \multimap A \leftrightarrow B$ is in G but not in H, because G' has no unshielded collider of B and C over A and $I(G) = I(G')$. □

Theorem 2. *A CG has the minimal set of bidirected edges for its Markov equivalence class if no feasible split is possible.*

Proof. Assume to the contrary there exists a CG G that does not have the minimal set of bidirected edges for its Markov equivalence class and no split is feasible. Hence there must exist a bidirected edge $X \leftrightarrow Y$ in G st $X \rightarrow Y$ exists in a CG G' st $I(G) = I(G')$. Let C be the component of X in G and U and L be two disjoint subsets of C st $C = U \cup L$, $X \in U$, $Y \in L$ and the subgraph of G induced by U is connected. It is trivial to see that such sets of nodes always must exist. If no split is feasible then we must have that one of the conditions in definition 1 fails for U and L. Hence one of the following assumptions must be true.

Assume there exists a node A st $A \in sp_G(U) \cap L$ for which $U \subseteq sp_G(A)$ does not hold. If this is the case there must exist an unshielded collider in G over a node D between A and E st $D \in U, D \in sp_G(A), E \in U, E \in sp_G(D), E \notin sp_G(A)$. This contradicts $I(G) = I(G')$ since $D \rightarrow A$ exists in G'.

Assume there exists a node A st $A \in sp_G(U) \cap L$ for which $pa_G(U) \subseteq pa_G(A)$ does not hold. If this is the case there must exist an unshielded collider in G over a node D between A and E st $D \in U, D \in sp_G(A), E \in pa_G(D), E \notin pa_G(A)$. This contradicts $I(G) = I(G')$ since $D \rightarrow A$ exists in G'.

Assume there exists a $B \in sp_G(L) \cap U$ for which $sp_G(B) \cap L$ is not complete. If this is the case there must exist an unshielded collider in G over a node B between D and E st $D, E \in L \cap sp_G(B), E \notin adj_G(D)$. This contradicts $I(G) = I(G')$ since $B \rightarrow D$ exists in G'.

This shows that if one of the constrains in definition 1 fails we can not have $I(G) = I(G')$ which contradicts the assumption. □

Finally, it is worth mentioning that we can guarantee that every member of a Markov equivalence class can be reached from any other member of that class by a sequence of feasible splits and mergings. The proof of this result can be seen at *http://www.ida.liu.se/~jospe/pgm12appendix.pdf*. This result is not used in this paper but we conjecture that it will play a central role in future work (see section 6).

4 Learning algorithm

In this section we present a constraint based algorithm which learns a CG from a probability distribution faithful to some CG. We then prove that the algorithm is correct and that the returned CG contains exactly the minimal set of bidirected edges for its Markov equivalence class.

The algorithm is very similar to the PC algorithm for directed acyclic graphs (Meek, 1995; Spirtes et al., 1993) and shares the same structure with the learning algorithms presented by both Studený for LWF CGs (Studený, 1997) and Peña for AMP CGs (Peña, 2012). The algorithm is shown in algorithm 1 and consists of four separate phases. In phase one (line 1-7) the adjacencies of the CG is recovered. In the second phase (line 8) the unshielded colliders are recovered. Phase three (line 9) then orients some of the remaining edges iff they are oriented in the same direction in all CGs G' st $I(G) = I(G')$. What remains for phase four (line 10-14) is then to orient the rest of the undirected edges st no new unshielded colliders or semi-directed cycles are introduced. This is done according to an algorithm presented in a proof by Koller and Friedman (2009, Theorem 4.13) and is possible since H_u is chordal as shown in Lemma 8.

The rules used in lines 8-9 in algorithm 1 are shown in figure 1. A rule is said to be applicable if the antecedent is satisfied for an induced subgraph of H. When a rule is applicable one of the non-arrow edge endings is then replaced with an arrow while the rest of the endings are kept the same. Which edge ending is orientated is shown in the consequent of each rule.

A rule is sound if the orientation introduced by the rule must be shared by every CG which contains the antecedent as an induced subgraph.

Lemma 3. *The rules 0 - 3 are sound.*

Proof. Rule 0: Since $B \notin S_{AC}$, $A \in adj_H(B)$ and $C \in adj_H(B)$ but $A \notin adj_H(C)$ we know that the following configurations can occur: $A \to B \leftarrow C$, $A \leftrightarrow B \leftarrow C$, $A \to B \leftrightarrow C$, $A \leftrightarrow B \leftrightarrow C$. In any other configuration, B would be in every separation set of A and C. In all these configurations we have $B \leftarrow\!\!\circ\ C$, so that edge must exist in G. Rule 1: If the edge was not directed in this direction we would have an unshielded collider of A and C over B that is not in G, because $B \in S_{AC}$. Rule 2: If we did not have the edge oriented in this direction we would have a semi-directed cycle in G. Rule 3: If the edge was orientated in the opposite direction, by applying rule 2 we would have an unshielded collider of B and C over A that is not in G, because

<u>Learning Algorithm</u>

Given a probability distribution p faithful to an unknown CG G, the algorithm learns a CG H st $I(H) = I(G)$. Moreover, H has exactly the minimum set of bidirected edges for its equivalence class.

1 Let H denote the complete undirected graph
2 For $l = 0$ to $l = |V_H| - 2$
3   Repeat while possible
4   Select any ordered pair of nodes A and B in H st $A \in adj_H(B)$ and $|adj(A) \setminus B| \geq l$
5   If there exists $S \subseteq (adj_H(A) \setminus B)$ st $|S| = l$ and $A \perp_p B | S$ then
6   Set $S_{AB} = S_{BA} = S$
7   Remove the edge $A - B$ from H
8 Apply rule 0 while possible
9 Apply rules 1-3 while possible
10 Let H_u be the subgraph of H containing only the nodes and the undirected edges in H
11 Let T be the clique tree of H_u
12 Order the cliques $C_1, ..., C_n$ of H_u st C_1 is the root of T and if C_i is closer to the root than C_j in T then $C_i < C_j$.
13 Order the nodes st if $A \in C_i$, $B \in C_j$ and $C_i < C_j$ then $A < B$
14 Orient the undirected edges in H according to the ordering obtained in line 15
15 Return H

Algorithm 1: Learning algorithm

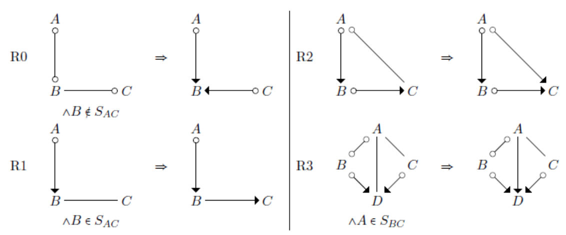

Figure 1: The rules

$A \in S_{BC}$. □

Lemma 4. *G and H have the same adjacencies after line 7.*

Proof. Assume to the contrary that there exists an adjacency between two nodes A and B in G st $B \notin de_G(A)$ that does not exist in H or vice versa. We know that, since G is faithful to p, all separation statements in G corresponds to independence statements in p and vice versa. We will first cover the case where the adjacency is in G but not in H. This means that there exists no separation set S st $A \perp_p B | S$ since $A \perp_G B | S$ does not hold for any S. Hence the prerequisite for line 5 in the algorithm can never be true and so the edge between A and B can never be removed and we have a contradiction.

Secondly we will cover if there is an adjacency in H but not in G. If there is no adjacency in G between two nodes A and B we know that there has to exist a separation set S st $A \perp_p B | S$ and that $S \subseteq pa_G(A)$ according to the definition of separation. We know, from the paragraph above, that all adjacencies in G must exist in H which means that $pa_G(A) \subseteq adj_H(A)$. However no such set was found in line 5, hence we have a contradiction. □

Lemma 5. *G and H have the same unshielded colliders and adjacencies after line 8.*

Proof. From Lemma 4 we know that G and H have the same adjacencies. First, assume to the contrary that there exists an unshielded collider in G but not in H after line 8. That means we have an unshielded collider between A and B over C st $C \notin S_{AB}$, $C \in adj_G(A)$ and $C \in adj_G(B)$ but $A \notin adj_G(B)$. According to Lemma 4 the same adjacencies must also hold in H. We know that if we do not have an unshielded collider in H we have $A \circ\!\!-\!\!\circ C \multimap B$. This does however fulfill the prerequisite for rule 0, and hence it should have been applied and we have a contradiction.

Secondly it follows from Lemma 3 that rule 0 is sound so there can be no unshielded colliders in H that are not in G after line 8. This brings us to a contradiction. □

Lemma 6. *After line 9, H cannot have a subgraph of the form $A \circ\!\!\rightarrow B - C$ without also having the edge $A \rightarrow C$.*

Proof. Assume the contrary. First we must have $A \in adj_H(C)$ or rule 1 would orient $B - C$[1]. We can see that H can not have the edge $A \leftarrow\!\!\circ C$ or rule 2 would be applicable for $B - C$. If H has the edge $A \rightarrow C$ we are done, which leaves us with $A - C$.

Secondly we will study why A and B can have an orientation $A \circ\!\!\rightarrow B$. This can be because of one of four reasons.

Case 1: Edge $A \circ\!\!\rightarrow B$ was orientated using rule 0. This means that there exists a node D st $D \circ\!\!\rightarrow B$ and $D \notin adj_H(A)$. $D \in adj_H(C)$ must hold or rule 1 would orient $B - C$[1], but then rule 3 is applicable for the edge $B - C$ so this can not be the reason.

Case 2: Edge $A \circ\!\!\rightarrow B$ was orientated using rule 1. This means that the edge $D \circ\!\!\rightarrow A$ st $D \notin adj(B)$ existed when $A \circ\!\!\rightarrow B$ was oriented. $D \in adj(C)$ must hold or the edge $A - C$ would be oriented by rule 1[1]. Restart the proof with $D \circ\!\!\rightarrow A - C$ and it can be seen that this configuration is impossible. Hence this case can not be the reason.

Case 3: Edge $A \circ\!\!\rightarrow B$ was orientated because of rule 3. This means that we have an unshielded collider of two nodes D and E st $D \circ\!\!\rightarrow B$,$E \circ\!\!\rightarrow B$ $A \in adj_H(D), A \in adj_H(E)$ and $D \notin adj_H(E)$. Now $D \in adj_H(C)$ and $E \in adj_H(C)$ must hold or rule 1 would orient $B - C$[1]. We know that there can be no unshielded collider over C between D and E, otherwise rule 3 would be applicable on $A - C$[1]. This gives us that we have $D - C$ and/or $E - C$, since if the edge orientated towards D or E rule 2 would be applicable. However, with this configuration rule 3 is applicable[1] for edge $B - C$ so this can not be the reason.

Case 4: Edge $A \circ\!\!\rightarrow B$ was directed using rule 2. This means that the edges $A \circ\!\!\rightarrow D$ and $D \circ\!\!\rightarrow B$ existed in H when $A \circ\!\!\rightarrow B$ was oriented. $D \in adj_H(C)$, otherwise rule 1 would orient the edge $B - C$[1]. $D - C$ must hold or rule 2 would cause an orientation of $B - C$ or $A - C$. Restart the proof with $A \circ\!\!\rightarrow D - C$ and it can be seen that this configuration is impossible, hence case 4 can not be the reason of the orientation. □

Lemma 7. *After line 9, H can not have a subgraph of the form* $A \leftrightarrow B - C$.

Proof. Assume the contrary. Then, according to Lemma 6, H also has the edge $A \rightarrow C$. If this is the case then rule 2 is applicable and we have a contradiction. □

Lemma 8. *After line 9, removing all directed edges and bidirected edges from H results in a chordal graph.*

Proof. Assume to the contrary that there exists a non-chordal cycle $V_1 - V_2 - ... - V_n - V_1$ and $n > 3$. We know that there exists a CG G that is faithful to the probability distribution p and has the same unshielded colliders and adjacencies as H by Lemma 5. From Lemma 3 we also know that the rules are sound, i.e. that the orientations in H are in G. Hence we know that there still must exist a valid way to orient the edges in H that remain undirected after line 9.

Now if we would orient an edge in the cycle st $V_1 \circ\!\!\rightarrow V_2 - ... - V_n - V_1$ it is easy to see that we would have to orient every other edge st $V_i \rightarrow V_{i+1}$ or we would have a new unshielded collider over some V_j since $V_{j-1} \notin adj_H(V_{j+1})$. If we would orient all edges in this direction we would however have a semi-directed cycle and hence there exists no possible orientation that results in a CG. □

Lemma 9. *In line 14, when undirected edges are orientated, no new unshielded colliders are introduced to H.*

Proof. It follows directly from Lemma 6 that no undirected edge can be orientated such that it creates an unshielded collider with an edge oriented by rules 0-3. Secondly it is proven (Koller and Friedman, 2009, Theorem 4.13) that no unshilded colliders are created between different undirected edges with the algorithm presented in lines 10-14. □

[1]Note that if we have $X \circ\!\!-\!\!\circ Y -\!\!\circ Z$ in H and $X \notin adj_H(Z)$ we must also have $Y \in S_{XY}$ or rule 0 would have been applicable in line 8.

Lemma 10. *No semi-directed cycle exists in H after line 15.*

Proof. Assume to the contrary, that a semi-directed cycle exists. If the semi-directed cycle is of length 3 then it could have been be created in one of the following ways:

Case 1: All its edges got oriented by rules 0-3. However, this is a contradiction because rule 2 would be applicable and no semi-directed cycle would remain after applying the rule.

Case 2: All its edges got oriented by line 14. However, this is a contradiction by Theorem 4.13 in Koller and Friedman (2009).

Case 3: Some edges got oriented by rules 0-3 and some by line 14. Then, after applying the rules, the cycle contained the configuration $A \rightarrow B - C$ or $A \leftrightarrow B - C$ for some nodes A, B, C in the cycle. The first one is impossible because, otherwise, $A \rightarrow C$ by Lemma 6 and, thus, there would not be semi-directed cycle of length 3, which contradicts the paragraph above. The latter one is impossible by Lemma 7.

Hence we can have no semi-directed cycle of length 3. Now assume the semi-directed cycle is of length 4. It could have been be created in one of the following ways:

Case 1: All its edges got oriented by rules 0-3. This implies that the semi-directed cycle in H is actually a cycle with only bidirected edges in G because the rules are sound. However, this implies that there is an unshielded collider in G that was not in H, which contradicts Lemma 5, or that H had a semi-directed cycle of length 3, which contradicts the paragraph above.

Case 2: All its edges got oriented by line 14. However, this is a contradiction by Theorem 4.13 in Koller and Friedman (2009).

Case 3: Some edges got oriented by rules 0-3 and some by line 14. Then, after applying the rules, the cycle contained the configuration $A \rightarrow B - C$ or $A \leftrightarrow B - C$. The first one is impossible because, otherwise, $A \rightarrow C$ by Lemma 6 and there would not be semi-directed cycle. The latter one is impossible by Lemma 7.

Hence H has no semi-directed cycle of length 4. Now, repeating the reasoning for semi-directed cycles of length 5, 6, etc. it is easy to see that no semi-directed cycle can exist since H has a bounded number of nodes. □

Theorem 3. *After line 15, H is a CG and* $I(H) = I(G)$.

Proof. Lemma 3, 5 and 9 gives that $I(H) = I(G)$ after line 14. Lemma 10 then refutes that H can contain a semi-directed cycle. □

Theorem 4. *After line 15, H has exactly the unique minimal set of bidirected edges for its Markov equivalence class.*

Proof. It is clear that bidirected edges only can be introduced to H by rules 0-3. From Lemma 3 it also follows that the rules are sound meaning that all the orientations in H caused by the rules have to exist in every G' st $I(G') = I(G)$. □

5 Conclusion

In this paper we have presented and proved two fundamental elements for the multivariate regression interpretation of CGs. The first element was an algorithm to learn a multivariate regression CG from a probability distribution faithful to some CG. The second element we have presented is a feasible split and a feasible merging st they alter the structure of a CG but do not change the Markov equivalence class of the CG. We have also shown that there exists a set of CGs with a unique minimal set of bidirected edges for each Markov equivalence class and that these CGs can be reached from any CG in the class using the split operation.

6 Further Work

Many elements are still not presented or proven for the multivariate regression interpretation of CGs. The natural continuation of the work presented here would be to develop a learning algorithm with weaker assumptions than the one presented. This could for example be a learning algorithm which only assumes that the probability distribution fullfills the composition property. A second natural continuation of the work here would be to use the split and merging operations to prove that Meek's conjecture holds for the multivariate regression interpretation of CGs, similar way to Peña's work within the LWF interpretation (Peña, 2011). This could then be used to develop learning algorithms that are correct under the composition property.

Acknowledgments

This work is funded by the Center for Industrial Information Technology (CENIIT) and a so-called career contract at Linköping University, and by the Swedish Research Council (ref. 2010-4808).

References

Steen A. Andersson, David Madigan and Michael D. Perlman. 2001. Alternative Markov Properties for Chain Graphs. *Scandinavian Journal of Statistics*, 28:33–85.

David R. Cox and Nanny Wermuth. 1993. Linear Dependencies Represented by Chain Graphs. *Statistical Science*, 8:204-283.

David R. Cox and Nanny Wermuth. 1996. *Multivariate Dependencies: Models, Analysis and Interpretation.* Chapman and Hall.

Mathias Drton. 2009. Discrete Chain Graph Models. *Bernoulli* 15(3):736–753.

Daphne Koller and Nir FriedMan. 2009. *Probabilistic Graphcal Models.* MIT Press.

Steffen L. Lauritzen. 1996. *Graphical Models.* Oxford University Press.

Zongming Ma, Xianchao Xie and Zhi Geng. 2008. Structural Learning of Chain Graphs via Decomposition. *Journal of Machine Learning Research*, 9:2847–2880.

Christopher Meek. 1995. Causal Inference and Causal Explanation with Background Knowledge. *Proceedings of 11th Conference on Uncertainty in Artificial Intelligence*, 403–418.

Jose M. Peña. 2011. Towards Optimal Learning of Chain Graphs. arXiv:1109.5404v1 [stat.ML].

Jose M. Peña. 2012. Learning AMP Chain Graphs under Faithfulness arXiv:1204.5357v1 [stat.ML].

Thomas S. Richardson. 2003. Markov Properties for Acyclic Directed Mixed Graphs. *Scandinavian Journal of Statistics*, 30(1):145–157.

Peter Spirtes, Clark Glymour and Richard Scheines. 1993. *Causation, Prediction, and Search.* Springer-Verlag.

Milan Studený. 1997. A Recovery Algorithm for Chain Graphs. *International Journal of Approximate Reasoning*, 17:265–293.

Milan Studený, Alberto Roverato and Šárka Štěpánová. 2009. Two Operations of Merging and Splitting Components in a Chain Graph. *Kybernetika*, 45:208–248.

Nanny Wermuth and Kayvan Sadeghi. 2012. Sequences of regressions and their independences. *Invited paper in TEST.* Available online.