

MARGINALIZATION AND CONDITIONING FOR LWF CHAIN GRAPHS¹

BY KAYVAN SADEGHI

University of Cambridge

In this paper, we deal with the problem of marginalization over and conditioning on two disjoint subsets of the node set of chain graphs (CGs) with the LWF Markov property. For this purpose, we define the class of chain mixed graphs (CMGs) with three types of edges and, for this class, provide a separation criterion under which the class of CMGs is stable under marginalization and conditioning and contains the class of LWF CGs as its subclass. We provide a method for generating such graphs after marginalization and conditioning for a given CMG or a given LWF CG. We then define and study the class of anteril graphs, which is also stable under marginalization and conditioning and contains LWF CGs, but has a simpler structure than CMGs.

1. Introduction. Graphical models use graphs, in which nodes are random variables and edges indicate some types of conditional dependencies. Mixed graphs, which are graphs with several types of edges, have started to play an important role in graphical models as they can deal with more complex independence structures that arise in different statistical studies.

The first example of mixed graphs in the literature appeared in [11]. This was a chain graph (CG) with a specific interpretation of conditional independence, which is now generally known as the Lauritzen–Wermuth–Frydenberg or LWF interpretation. A formal interpretation, that is, a Markov property, was later provided by [5]. This Markov property, together with other properties such as the factorization property was extensively discussed in [9]. By the term LWF CGs, one refers to the class of CGs with a specific independence structure that comes from the LWF Markov property.

It has become apparent that CGs with the LWF interpretation of independencies are important tools in capturing conditional independence structure of various probability distributions. For example, Studený and Bouckaert [24] showed that for every CG, there exists a strictly positive discrete probability distribution that embodies exactly the independence statements displayed by the graph, and Peña [15] proved that almost all the regular Gaussian distributions that factorize with

Received May 2014; revised January 2016.

¹Supported by Grant #FA9550-12-1-0392 from the U.S. Air Force Office of Scientific Research (AFOSR) and the Defense Advanced Research Projects Agency (DARPA).

MSC2010 subject classifications. Primary 62H99; secondary 62A99.

Key words and phrases. c -separation criterion, chain graph, independence model, LWF Markov property, m -separation, marginalization and conditioning, mixed graph.

respect to a chain graph are faithful to it. This means that a Gaussian distribution chosen at random to factorize as specified by the LWF CG will have the independence structure of the graph and will satisfy no more independence constraints.

However, in the corresponding models to LWF CGs, when some variables are unobserved—also called latent or hidden—or when some variables are set to specific values, the implied independence structure, that is, the corresponding independence structure after marginalization and conditioning respectively, is not well understood.

The same problem for the well-known class of directed acyclic graphs (DAGs), which is a subclass of LWF CGs, has been a subject of study, and several classes of graphs have been defined in order to capture the marginal and conditional independence structure of DAGs. These include MC graphs [8], ancestral graphs [18] and summary graphs [26]; see also [19]. There is also a literature pertaining to this problem for other types of graphs; see, for example, the class of marginal AMP chain graphs in [16] for marginalization in AMP chain graphs [1].

For LWF CGs, as it will be shown in this paper, one can capture the independence structure induced by conditioning on some variables by another LWF CG, but in general cannot capture the independence structure induced by marginalization over some variables by a CG. In this sense, CGs are stable under conditioning but not under marginalization.

Indeed models with latent variables do not necessarily possess the desirable statistical properties of graphical models without latent variables, such as identifiability, existence of a unique MLE, or being curved exponential families in some cases such as DAGs; see, for example, [6].

However, a first step in dealing with this problem is, in the case of marginalization, to come up with a more complex class of graphs with a certain independence interpretation that captures the marginal independence structure of CGs; and in both cases of marginalization and conditioning, to provide methods by which the graphs that capture the marginal and conditional independence structure are generated. These are the main objectives of the current paper.

In the causal language (see, e.g., [13]) the resulting classes of graphs give a simultaneous representation to “direct effects”, “confounding”, and “non-causal symmetric dependence structures”.

It is important to note that the classes of graphs introduced here only deals with the conditional independence constraints, and not other constraints such as so-called *Verma constraints* [25]. The actual statistical model is much more complicated even when marginalizing DAGs; see, for example, [21].

The introduction of these classes of graphs is also justified in the paper by showing that, for large subclasses of these classes of graphs, there are probability distributions (in fact both Gaussian and discrete) that are faithful to them. Although finding the explicit parametrizations for the defined graphs is beyond the scope of this paper, it also seems possible to extend the existing parametrizations for

smaller types of graph in the literature to these classes in a fairly natural way. We will provide a discussion on this in the paper.

The structure of the paper is as follows: in the next section, we define mixed and chain graphs, and, for these classes of graphs, give graph theoretical definitions needed in this paper. In Section 3, we provide two equivalent ways for reading off independencies from a CG based on the LWF Markov property. In Section 4, we define the class of chain mixed graphs with certain independence interpretation, and show that they capture the marginal independence structure of LWF CGs and that they are stable under marginalization, and provide an algorithm for generating such graphs after marginalization. In Section 5, we show that the class of CMGs is also stable under conditioning, provide the corresponding algorithm, and combine marginalization and conditioning for CMGs. As a corollary, we see that LWF CGs are stable under conditioning. In Section 6, we define the class of anterital graphs as a subclass of CMGs, which also contains LWF CGs, and show that this class is stable under marginalization and conditioning. We also provide an algorithm for marginalization and conditioning for this class. In Section 7, we discuss the implications of the results for probabilistic independence models that are faithful to LWF CGs, and possible ways to generalize the parametrizations existing in the literature for CMGs and anterital graphs. In the Appendix in the supplementary material [20], we provide proofs of non-trivial lemmas, propositions and theorems in the paper as well as some more technical and yet less informative lemmas that are used in the proofs.

2. Definitions for mixed graphs and chain graphs.

2.1. Basic graph theoretical definitions. A *graph* G is a triple consisting of a *node* set or *vertex* set V , an *edge* set E , and a relation that with each edge associates two nodes (not necessarily distinct), called its *endpoints*. When nodes i and j are the endpoints of an edge, these are *adjacent* and we write $i \sim j$. We say the edge is *between* its two endpoints. We usually refer to a graph as an ordered pair $G = (V, E)$. Graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are called *equal* if $(V_1, E_1) = (V_2, E_2)$. In this case, we write $G_1 = G_2$.

Notice that graphs that we use in this paper (and in general in the context of graphical models) are so-called *labeled graphs*, that is, every node is considered a different object. Hence, for example, graph $i — j — k$ is not equal to $j — i — k$.

Here, we introduce some basic graph theoretical definitions. A *loop* is an edge whose endpoints are equal. *Multiple edges* are edges whose endpoints are the same as each other. A *simple graph* has neither loops nor multiple edges. A *complete graph* is a simple graph with all pairs of nodes adjacent.

A *subgraph* of a graph G_1 is graph G_2 such that $V(G_2) \subseteq V(G_1)$ and $E(G_2) \subseteq E(G_1)$ and the assignment of endpoints to edges in G_2 is the same as in G_1 . An

induced subgraph by a subset A of the node set is a subgraph that contains the node set A and all edges between two nodes in A .

A *walk* is a list $\langle i_0, e_1, i_1, \dots, e_n, i_n \rangle$ of nodes and edges such that for $1 \leq m \leq n$, the edge e_m has endpoints i_{m-1} and i_m . A *path* is a walk with no repeated node or edge. A *cycle* is a walk with no repeated node or edge except $i_0 = i_n$. If the graph is simple then a path or a cycle can be determined uniquely by an ordered sequence of nodes. Throughout this paper, however, we use node sequences to describe paths and cycles even in graphs with multiple edges, but we assume that the edges of the path are all determined. It is usually apparent from the context or the type of the path which edge belongs to the path in multiple edges. We say a walk or a path is *between* the first and the last nodes of the list in G . We call the first and the last nodes *endpoints* of the walk or of the path. All other nodes are the *inner nodes*.

For a walk or path $\pi = \langle i_1, \dots, i_n \rangle$, any subsequence $\langle i_k, i_{k+1}, \dots, i_{k+p} \rangle$, $1 \leq k, k+p \leq n$, whose members appear consecutively on π , defines a *subwalk* or a *subpath* of π , respectively.

2.2. Some definitions for mixed graphs. A *mixed graph* is a graph containing three types of edges denoted by arrows, arcs (two-headed arrows), and lines (solid lines). Mixed graphs may have multiple edges of different types but do not have multiple edges of the same type. We do not distinguish between $i \longrightarrow j$ and $j \longrightarrow i$ or $i \longleftrightarrow j$ and $j \longleftrightarrow i$, but we do distinguish between $j \longrightarrow i$ and $i \longrightarrow j$. In this paper, we are only considering mixed graphs that do not contain loops of any type. These constitute the class of *loopless mixed graphs*.

For mixed graphs, we say that i is a *neighbor* of j if these are endpoints of a line, and i is a parent of j and j is a child of i if there is an arrow from i to j . We also define that i is a *spouse* of j if these are endpoints of an arc. We use the notation $\text{ne}(j)$, $\text{pa}(j)$, and $\text{sp}(j)$ for the set of all neighbors, parents, and spouses of j , respectively.

In the cases of $i \longrightarrow j$ or $i \longleftrightarrow j$, we say that there is an arrowhead pointing to (at) j .

A walk $\langle i = i_0, i_1, \dots, i_n = j \rangle$ is *directed* from i to j if all $i_k i_{k+1}$ edges are arrows pointing from i_k to i_{k+1} . If there is a directed walk from j to i then j is an *ancestor* of i and i is a *descendant* of j . We denote the set of ancestors of i by $\text{an}(i)$. Notice that, unlike some authors, we do not consider i to be in the set of ancestors or descendants of i . Moreover, a cycle with the above property is called a *directed cycle*.

A walk $\langle i = i_0, i_1, \dots, i_n = j \rangle$ from i to j is a *semi-directed walk* if it only consists of lines and arrows (it may contain only one type of edge), and every arrow $i_k i_{k+1}$ is pointing from i_k to i_{k+1} . Thus, a directed walk is a type of semi-directed walk. We shall say that i is *anterior* of j if there is a semi-directed walk from i to j . We use the notation $\text{ant}(i)$ for the set of all antecedents of i . Notice again that, similar to ancestors, we do not consider a node i to be an anterior of itself. For a set of nodes A , we define $\text{ant}(A) = \bigcup_{i \in A} \text{ant}(i) \setminus A$. Notice also that, since

ancestral graphs have no arrowheads pointing to lines, our definition of anterior extends the notion of anterior used in [18] for ancestral graphs. Moreover, a cycle with the properties of semi-directed walks is called a *semi-directed cycle*.

A *section* of a walk in a mixed graph is a maximal subwalk that only consists of lines. Thus, any walk decomposes uniquely into sections (that are not necessarily edge-disjoint and may also be single nodes). Similar to nodes, all sections on a walk between i and j are *inner sections* except those that contain i or j , which are *endpoint sections*. As in any walk, we can also define the endpoints of a section. A section ρ on a walk π is called a *collider* section if one of the three following walks is a subwalk of π : $i \rightarrow \rho \leftarrow j$, $i \leftarrow \rho \leftarrow j$, and $i \leftarrow \rho \rightarrow j$. All other sections on π are called *non-collider* sections. We may speak of collider or non-collider sections without mentioning the relevant walk when this is apparent from context.

A *trislide* on a walk π is a subpath $\langle i = i_0, i_1, \dots, i_n = j \rangle$, where ii_1 and $i_{n-1}j$ are arrows or arcs and the subpath $\rho' = \langle i_1, \dots, i_{n-1} \rangle$ is a section.

Three types of trislides $i \rightarrow o \longrightarrow \dots \longrightarrow o \leftarrow j$, $i \leftrightarrow o \longrightarrow \dots \longrightarrow o \leftarrow j$, and $i \leftrightarrow o \longrightarrow \dots \longrightarrow o \leftrightarrow j$ are *collider* trislides and all other types of trislides are *non-collider* on any walk of which the trislide is defined.

A *tripath* is a trislide where the subpath ρ' is a single node. Note that [19] used the term V-configuration for such a path. ([7] and most texts let a V-configuration be a tripath with non-adjacent endpoints.) Tripaths and their inner nodes can be defined to be colliders or non-colliders as trislides and their inner sections.

Two walks π_1 and π_2 (including trislides, tripaths or edges) between i and j are called *endpoint-identical* if there is an arrowhead pointing to the endpoint section containing i on π_1 if and only if there is an arrowhead pointing to the endpoint section containing i on π_2 ; and similarly for j . For example, the paths $i \rightarrow j$, $i \longrightarrow k \longrightarrow l \leftarrow j$ and $i \rightarrow k \leftrightarrow l \longrightarrow j$ are all endpoint-identical as they have an arrowhead pointing to the section containing j but no arrowhead pointing to the section containing i on the paths, but they are not endpoint-identical to $i \longrightarrow k \leftrightarrow j$.

2.3. Chain graphs. *Chain graphs* (CGs) consist of lines and arrows and do not contain any semi-directed cycles with at least one arrow.

It is implied from the definition that CGs are characterized by having a node set that can be partitioned into disjoint subsets forming so-called *chain components*. These are connected subgraphs consisting only of undirected edges and are obtained by removing all arrows in the graph. All edges between nodes in the same chain component are lines, and all edges between different chain components are arrows. In addition, the chain components can be ordered in such a way that all arrows point from a chain with a higher number to one with a lower number.

For example, in Figure 1(a) the graph is a chain graph with chain components $\tau_1 = \{l, j, k\}$, $\tau_2 = \{h, q\}$, and $\tau_3 = \{p\}$, but in Figure 1(b) the graph is not a chain graph because of the existence of the $\langle h, k, q \rangle$ semi-directed cycle. If one replaces

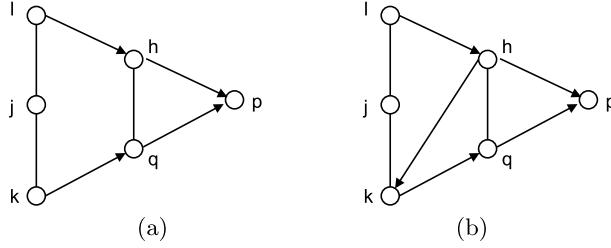


FIG. 1. (a) A CG. (b) A mixed graph that is not a CG.

every chain component with a single node, one obtains a *directed acyclic graph* (DAG), a graph consisting exclusively of arrows and without any directed cycles.

Notice that generally CGs are defined to contain arrows and one symmetric type of edge in their chain components (namely lines or arcs). In this sense, the type of CG with which we deal in this paper is a *line CG*.

3. LWF Markov property for CGs. An *independence model* \mathcal{J} over a set V is a set of triples $\langle X, Y | Z \rangle$ (called *independence statements*), where X , Y , and Z are disjoint subsets of V and Z can be empty, and $\langle \emptyset, Y | Z \rangle$ and $\langle X, \emptyset | Z \rangle$ are always included in \mathcal{J} . The independence statement $\langle X, Y | Z \rangle$ is interpreted as “ X is independent of Y given Z ”. Notice that independence models contain probabilistic independence models as a special case. For further discussion on independence models, see [23].

A graph G also induces an independence model $\mathcal{J}(G)$. One way is by using a *separation criterion*, which determines whether for three disjoint subsets A , B , and C of the node set of G , $\langle A, B | C \rangle \in \mathcal{J}(G)$. Such a criterion verifies whether A is separated from B by C in the sense that there are no walks or paths of specific types between A and B given C in the graph. Such a separation is denoted by $A \perp B | C$. It is clear that $\mathcal{J}(G)$ satisfies the *global Markov property*, which states that if $A \perp B | C$ in G then $\langle A, B | C \rangle \in \mathcal{J}$.

For CGs, at least four different separation criteria, that is, four different types of global Markov property have been discussed in the literature. Drton [3] has classified them as (1) the *LWF* or *block concentration* Markov property, (2) the *AMP* or *concentration regression* Markov property, as defined and studied by [1], (3) a Markov property that is dual to the AMP Markov property and (4) the *multivariate regression* Markov property, as introduced by [2] and studied extensively recently; for example, see [12, 27].

In this paper, we are interested in the LWF Markov property, and we introduce two equivalent separation criteria for this in this section. Henceforth, for the sake of brevity, by CGs we refer to CGs with the LWF Markov property.

The *moralization criterion* for CGs was defined in [5] and is a generalization of the moralization criterion for DAGs defined in [10]; see also [9]. The *moral graph*

of a chain graph G , denoted by $(G)^m$ is a graph that consists only of lines and that is generated from G as follows: for every edge ij in G there is a line ij in $(G)^m$. In addition, if nodes i and j are parents of the same chain component in G then there is the line ij in $(G)^m$.

Now let $G_{\text{ant}(A \cup B \cup C)}$ be the induced subgraph of G generated by $\text{ant}(A \cup B \cup C)$. The moralization criterion states that for A , B and C , three disjoint subsets of the node set of G , if there are no paths between A and B in $(G_{\text{ant}(A \cup B \cup C)})^m$ whose inner nodes are outside C then $A \perp_{\text{mor}} B \mid C$.

An equivalent criterion, called the *c-separation criterion* for CGs was defined in [24]. Here, we present a simpler version of that criterion, presented in [22], with a different notation and wording:

A walk π in a CG is a *c-connecting walk* given C if every collider section of π has a node in C and all non-collider sections are outside C . A section on π is *open* if either: it is a collider section and one of its nodes is in C ; or it is a non-collider section and all its nodes are outside C . Otherwise, it is *blocked*. We say that A and B are *c-separated* given C if there are no *c-connecting walks* between A and B given C , and we use the notation $A \perp_c B \mid C$.

Notice that, as mentioned in [24], there is potentially an infinite number of walks and, therefore, this might not be an appropriate criterion for testing independencies. Although, in this paper, we only use this criterion in order to prove our theoretical results regarding marginalization and conditioning, and an infinite number of walks is not an issue for this purpose, in [22], it was shown that this criterion can also be implemented with an algorithm.

For example, in the graph of Figure 2(a), the independence statement $j \perp h \mid l$ does not hold. This can be seen by looking at the moral graph $(G_{\text{ant}(\{j,h,l\})})^m = (G_{\{j,h,k,q,l,r\}})^m$ in Figure 2(b), and observing that the inner nodes of the path $\langle j, k, q, h \rangle$ are outside the conditioning set. The same conclusion can be made by looking at the walk $\langle j, k, l, r, q, h \rangle$, where the non-collider sections k and q are outside the conditioning set, but the inner node l of the collider section $\langle l, r \rangle$ is in the conditioning set.

The equivalence of the moralization criterion and the original *c-separation criterion* was proven in Consequence 4.1 in [24]. The equivalence with the mentioned simplified criterion was proven in [22]. We use the notation $\mathcal{J}_c(G)$ for the independence model induced from G by the above criteria.

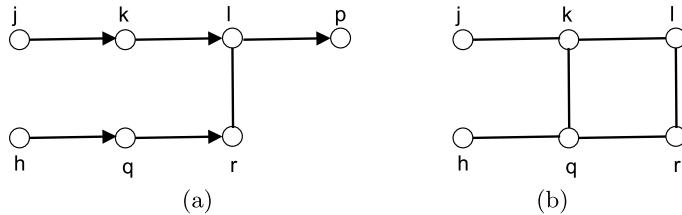


FIG. 2. (a) A chain graph G . (b) The moral graph $(G_{\text{ant}(\{j,h,l\})})^m$.

We first prove the following lemma, which provides an equivalent type of walk to c -connecting walks.

LEMMA 1. *There is a c -connecting walk between i and j given C if and only if there is a walk between i and j whose sections are all paths, and on which nodes of every collider section are in $C \cup \text{ant}(C)$, and non-collider sections are outside C . In addition, these walks can be chosen to be endpoint-identical.*

Notice that by the same method as the proof of this lemma, one can always assume that a section on a walk is a path. This is our assumption throughout the paper unless otherwise stated.

4. Stability of CGs under marginalization and conditioning. For a subset C of V , the *independence model after conditioning on C* , denoted by $\alpha(\mathcal{J}; \emptyset, C)$, is

$$\alpha(\mathcal{J}; \emptyset, C) = \{\langle A, B | D \rangle : \langle A, B | D \cup C \rangle \in \mathcal{J} \text{ and } (A \cup B \cup D) \cap C = \emptyset\}.$$

One can observe that $\alpha(\mathcal{J}; \emptyset, C)$ is an independence model over $V \setminus C$.

We now present the definition of *stability under conditioning* [19]: Consider a family of graphs \mathcal{T} . If, for every graph $G = (V, E) \in \mathcal{T}$ and every disjoint subsets C of V , there is a graph $H \in \mathcal{T}$ such that $\mathcal{J}(H) = \alpha(\mathcal{J}(G); \emptyset, C)$ then \mathcal{T} is stable under conditioning. Notice that the node set of H is $V \setminus C$.

We will see as a corollary of the results and algorithms in the next section that CGs are stable under conditioning.

Similar to the conditioning case, for a subset M of V , the *independence model after marginalization over M* , denoted by $\alpha(\mathcal{J}; M, \emptyset)$, is defined by

$$\alpha(\mathcal{J}; M, \emptyset) = \{\langle A, B | D \rangle \in \mathcal{J} : (A \cup B \cup D) \cap M = \emptyset\}.$$

One can observe that $\alpha(\mathcal{J}; M, \emptyset)$ is an independence model over $V \setminus M$.

The definition of *stability under marginalization* is defined similarly to the conditioning case: for a family of graphs \mathcal{T} , if, for every graph $G = (V, E) \in \mathcal{T}$ and every disjoint subsets C of V , there is a graph $H \in \mathcal{T}$ such that $\mathcal{J}(H) = \alpha(\mathcal{J}(G); M, \emptyset)$ then \mathcal{T} is stable under marginalization. We see again that the node set of H is $N = V \setminus M$.

CGs are not closed under marginalization. For example, it can be shown that G in Figure 3 is a CG (in fact a DAG) whose induced marginal independence model cannot be represented by a CG. We leave the details as an exercise to the reader.

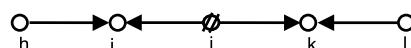


FIG. 3. A chain graph G , by which it can be shown that the class of CGs is not stable under marginalization. ($\emptyset \in M$.)

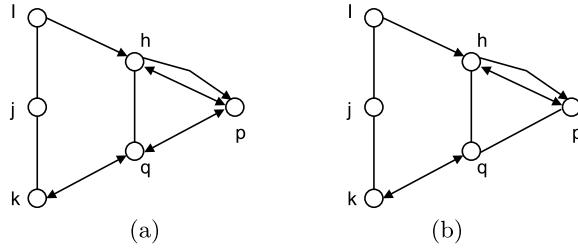


FIG. 4. (a) A CMG. (b) A mixed graph that is not a CMG.

Hence, we define a class of graphs that is stable under marginalization and contains CGs: the class of *chain mixed graphs* (CMGs) is the class of mixed graphs without semi-directed cycles with at least an arrow. Notice that we allow CMGs to have multiple edges consisting of arcs and arrows and arcs and lines. This is a generalization of chain graphs since if a CMG does not contain arcs then it is a chain graph.

For example, in Figure 4(a) the graph is a CMG, but in Figure 4(b) the graph is not a CMG because of the existence of the $\langle h, p, q \rangle$ semi-directed cycle.

We provide a c -separation criterion for CMGs, and using this, show that CMGs are closed under marginalization. For this purpose, we provide in this section an algorithm that, from a CMG (or a chain graph) G and after marginalization over M , generates a CMG with the corresponding independence model after marginalization over M .

We define a c -separation criterion for CMGs with exactly the same wordings as that of CGs: a walk π in a CG is a c -connecting walk given C if every collider section of π has a node in C and all non-collider sections are outside C . We say that A and B are c -separated given C if there are no c -connecting walks between A and B given C , and we use the notation $A \perp_c B | C$.

However, notice that this is in fact a generalization of the c -separation criterion for CGs since, for CMGs, bidirected edges on π may make a section collider.

We now provide an algorithm that, from a chain mixed graph G and after marginalization over M , generates a CMG with the corresponding independence model after marginalization over M . Notice that this algorithm may indeed be applied to a CG.

ALGORITHM 1 [$\alpha_{\text{CMG}}(G; M, \emptyset)$: Generating a CMG from a chain mixed graph G after marginalization over M].

Start from G .

1. Generate an ij edge as in Table 1, steps 8 and 9, between i and j on a collider trislide with an endpoint j and an endpoint in M if the edge of the same type does not already exist.

TABLE 1
*Types of edge induced by tripaths with inner node $m \in M$
and trislides with endpoint $m \in M$*

1	$i \leftarrow m \leftarrow j$	generates	$i \leftarrow j$
2	$i \leftarrow m \longrightarrow j$	generates	$i \leftarrow j$
3	$i \longleftarrow m \longrightarrow j$	generates	$i \longleftrightarrow j$
4	$i \leftarrow m \rightarrow j$	generates	$i \longleftrightarrow j$
5	$i \leftarrow m \longleftrightarrow j$	generates	$i \longleftrightarrow j$
6	$i \longrightarrow m \leftarrow j$	generates	$i \leftarrow j$
7	$i \longrightarrow m \longrightarrow j$	generates	$i \longrightarrow j$
8	$m \rightarrow i \longrightarrow \dots \longrightarrow o \leftarrow j$	generates	$i \leftarrow j$
9	$m \rightarrow i \longrightarrow \dots \longrightarrow o \longleftrightarrow j$	generates	$i \longleftrightarrow j$

2. Generate an appropriate edge as in Table 1, steps 1 to 7, between the endpoints of every tripath with inner node in M if the edge of the same type does not already exist. Apply this step until no other edge can be generated.
3. Remove all nodes in M .

Notice that, here and elsewhere, by removing nodes we mean also removing all the adjacent edges to those nodes. Notice also that all the cases generate an endpoint-identical edge to the tripath or the trislide. In addition, in cases 8 and 9, the node m is separate from the inner nodes of the concerned trislide since otherwise there will be a semi-directed cycle in the graph.

This algorithm is a generalization of the marginalization part of the summary-graph-generating algorithm [19]. The first seven cases are exactly the same as the corresponding cases in the summary-graph-generating algorithm, whereas cases 8 and 9 do not appear in the summary-graph-generating algorithm since in summary graphs there are no arrowheads pointing to lines. The other reason is that here we deal with connecting walks instead of paths, and the subwalk $\langle i, m, i \rangle$ may be present in a connecting walk. In general, here in this algorithm, and in later algorithms in this paper, the sections are treated in the same way as the nodes are treated in the algorithms that generate summary graphs, acyclic directed mixed graphs (ADMGs) [17], or ancestral graphs. It is also worth noticing that all these algorithms are indeed generalizations of the ordinary latent projection operation; see [13].

Figure 5 illustrates how to apply Algorithm 1 step by step to a CG.

We consider Algorithm 1 a function denoted by α_{CMG} . Notice that for every chain mixed graph G , it holds that $\alpha_{\text{CMG}}(G; \emptyset, \emptyset) = G$. We first show that $\alpha_{\text{CMG}}(G; M, \emptyset)$ is a CMG.

PROPOSITION 1. *Graphs generated by Algorithm 1 are CMGs.*

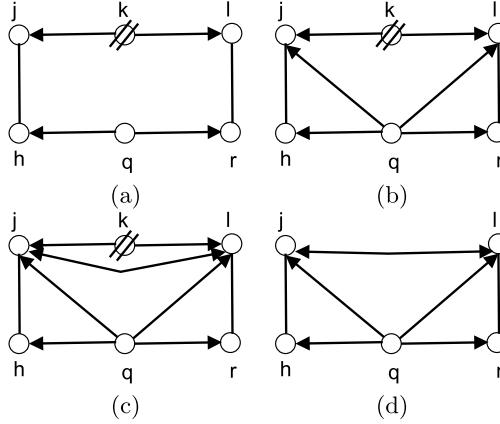


FIG. 5. (a) A chain graph G , $\emptyset \in M$. (b) The graph after applying step 1 of Algorithm 1 (case 8 of Table 1). (c) The graph after applying step 2 of Algorithm 1 (case 4 of Table 1). (d) The generated CMG after applying step 3.

We first provide lemmas that express the global behavior of step 2 of Algorithm 1 as well as a generalization and an implication of step 1 (in the Appendix in [20]).

LEMMA 2. *Let G be a CMG. There exists an edge between i and j in $\alpha_{\text{CMG}}(G; M, \emptyset)$ if and only if there exists an endpoint-identical walk between i and j in the graph generated after applying step 1 of Algorithm 1 to G whose inner sections are all non-collider and whose inner nodes are all in M .*

The following theorem shows that $\alpha_{\text{CMG}}(\cdot, \cdot, \emptyset)$ is well-defined in the sense that, instead of directly generating a CMG, we can split the nodes that we marginalize over into two parts, first generate the CMG related to the first part, then from the generated CMG, generate the desired CMG related to the second part.

THEOREM 1. *For a chain mixed graph G and disjoint subsets M and M_1 of its node set,*

$$\alpha_{\text{CMG}}(\alpha_{\text{CMG}}(G; M, \emptyset); M_1, \emptyset) = \alpha_{\text{CMG}}(G; M \cup M_1, \emptyset).$$

Some CMGs may not be generated after marginalization for CGs. In the following proposition, we provide the exact set of graphs to which CMGs are mapped after marginalization. Denote by \mathcal{CG} the set of all CGs and by \mathcal{CMG} the set of all CMGs.

PROPOSITION 2. *Define \mathcal{H} to be the subset of \mathcal{CMG} with the following properties:*

1. There is no collider trislide of form $k \longleftrightarrow i — \dots — j \leftarrow l$ unless there is an arrow from l to i ;
2. there is no collider trislide of form $k \longleftrightarrow i — \dots — j \longleftrightarrow l$ unless there are kj , il and ij arcs.

Then $\alpha_{\text{CMG}}(\cdot; \cdot, \emptyset)$ maps \mathcal{CG} and a subset of the node set of its member surjectively onto \mathcal{H} .

Here, we prove the main result of this section:

THEOREM 2. *For a chain mixed graph G and disjoint subsets A , B , M and C_1 of its node set,*

$$\langle A, B | C_1 \rangle \in \mathcal{J}_c(\alpha_{\text{CMG}}(G; M, \emptyset)) \iff \langle A, B | C_1 \rangle \in \mathcal{J}_c(G).$$

We, therefore, have the following immediate corollary.

COROLLARY 1. *The class of chain mixed graphs, \mathcal{CMG} , with c-separation criterion is stable under marginalization.*

5. Stability of CMGs under marginalization and conditioning.

5.1. Stability of CMGs under conditioning. In the previous section, we showed that the class of CMGs is stable under marginalization. In this section, we first show that the class of CMGs is also stable under conditioning, and provide an algorithm for conditioning for CMGs:

ALGORITHM 2 [$\alpha_{\text{CMG}}(G; \emptyset, C)$: Generating a CMG from a chain mixed graph G after conditioning on C].

Start from G .

1. Find all nodes in $C \cup \text{ant}(C)$ and call this set S .
2. For collider trislides illustrated in Table 2, steps 4 and 5, with an endpoint i and one endpoint in S , generate an ij edge following the table if the edge does not already exist.
3. For collider trislides (including tripaths) illustrated in Table 2, steps 1–3, with at least one inner node in S , generate an edge following the table if the edge does not already exist. Apply this step repeatedly until no other edge can be generated, but do not use generated lines (to generate new sections).
4. Remove the arrowheads of all arrows and arcs pointing to members of S (i.e., turn such arrows into lines and such arcs into arrows).
5. Remove all nodes in C .

TABLE 2
Types of edges induced by trislide with an inner node or endpoint $s \in S = C \cup \text{ant}(C)$

1	$i \rightarrow s \text{ --- } \dots \text{ --- } s \leftarrow j$	generates	$i \text{ --- } j$
2	$i \leftarrow s \text{ --- } \dots \text{ --- } s \leftarrow j$	generates	$i \leftarrow j$
3	$i \leftarrow s \text{ --- } \dots \text{ --- } s \rightarrow j$	generates	$i \rightarrow j$
4	$s \leftarrow i \text{ --- } \dots \text{ --- } \circ \leftarrow j$	generates	$i \leftarrow j$
5	$s \leftarrow i \text{ --- } \dots \text{ --- } \circ \rightarrow j$	generates	$i \rightarrow j$

Notice that if a node of a section is in S then all the inner nodes are in S , thus, we may speak of a section being in S . Notice also that all the steps of the algorithm generate endpoint-identical edges to the concerned trislide. In addition, we can assume that the endpoints of trislide are disjoint from the inner nodes, since (1) j as an endpoint of an arrow cannot be also an inner node because the graph does not contain semi-directed cycles; and (2) cases 2 and 3 with i an inner node are equivalent to cases 4 and 5, respectively, and cases 4 and 5 with s an inner node are equivalent to cases 2 and 3, respectively.

Similar to Algorithm 1, this algorithm is a generalization of the conditioning part of the summary-graph-generating algorithm [19]. The first three cases are the same when one considers sections here to be the nodes in the summary-graph-generating algorithm. Cases 4 and 5 do not appear in the summary-graph-generating algorithm for the same reasons explained before.

Figure 6 illustrates how to apply Algorithm 2 step by step to a CMG. First, let us provide a global interpretation of step 3 of Algorithm 2.

LEMMA 3. *Let G be a CMG. There exists an edge between i and j in the graph generated after step 3 of Algorithm 2 if and only if there exists an endpoint-identical walk to the edge between i and j in the generated graph after step 2 whose inner sections are all collider and in $C \cup \text{ant}(C)$, and whose endpoint sections contain a single node (i or j).*

We provide two lemmas that explain why the set S can be fixed in the beginning of the algorithm, and why there is no need to apply step 4 of Algorithm 2 repeatedly.

LEMMA 4. *Let G be a CMG. If there is an arrow from j to i or a line between j and i generated by steps 3 or 4 of Algorithm 2 then $j \in S = C \cup \text{ant}(C)$. In addition, generated lines by Algorithm 2 do not lie on any collider section in $\alpha_{\text{CMG}}(G; \emptyset, C)$.*

LEMMA 5. *Let G be a CMG. A node i is in $\text{ant}(C)$ in G if and only if it is in $\text{ant}(C)$ in the graph generated after every step of Algorithm 2 before step 5.*

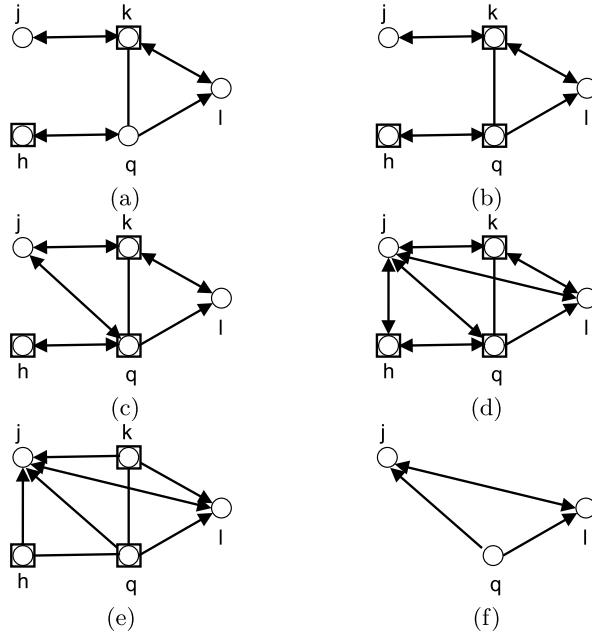


FIG. 6. (a) A chain mixed graph G , $\square \in C$. (b) The graph after applying step 1 of Algorithm 2, $\square \in S = C \cup \text{ant}(C)$. (c) The generated graph after applying step 2 (step 5 of Table 2). (d) The generated graph after applying step 3 (steps 2 and 3 of Table 2). (e) The generated graph after applying step 4. (f) The generated CMG from G .

We now follow the same procedure as in the previous section.

PROPOSITION 3. *Graphs generated by Algorithm 2 are CMGs.*

Here, we provide the global interpretation of Algorithm 2.

LEMMA 6. *Let G be a CMG. There exists an edge between i and j in $\alpha_{\text{CMG}}(G; \emptyset, C)$ if and only if there exists a walk between i and j in G whose inner sections are all collider and in $S = C \cup \text{ant}(C)$, and whose endpoint sections contain a single node (i or j) except when there is an arrowhead at the section containing i (or j), and i (or j) is a spouse of a member of S . In addition, the walk and the edge are endpoint-identical except when there is an arrowhead at the endpoint section containing i (or j), and $i \in \text{ant}(C)$ [or $j \in \text{ant}(C)$] in G .*

THEOREM 3. *For a chain mixed graph G and disjoint subsets C and C_1 of its node set,*

$$\alpha_{\text{CMG}}(\alpha_{\text{CMG}}(G; \emptyset, C); \emptyset, C_1) = \alpha_{\text{CMG}}(G; \emptyset, C \cup C_1).$$

THEOREM 4. *For a chain mixed graph G and disjoint subsets A, B, C and C_1 of its node set,*

$$\langle A, B \mid C_1 \rangle \in \mathcal{J}_c(\alpha_{\text{CMG}}(G; \emptyset, C)) \iff \langle A, B \mid C \cup C_1 \rangle \in \mathcal{J}_c(G).$$

COROLLARY 2. *The class of chain mixed graphs, \mathcal{CMG} , with c-separation criterion is stable under conditioning.*

Applying Algorithm 2 to a CG, step 2 becomes inapplicable, and step 3 specializes to generating a line between the endpoints of collider trislide with at least one inner node in S if the line does not already exist. Denote this specialization by $\alpha_{\text{CG}}(G, \emptyset, C)$. We first have the following.

PROPOSITION 4. *Algorithm 2 generates CGs from CGs.*

Denote now by \mathcal{CG} the set of all CGs. We also provide the following trivial statement.

PROPOSITION 5. *The map $\alpha_{\text{CG}}(\cdot; \emptyset, \cdot)$ from \mathcal{CG} and a subset of the node set of its members to \mathcal{CG} is surjective.*

PROOF. The result follows from the fact that $\alpha_{\text{CMG}}(G; \emptyset, \emptyset) = G$. \square

We, therefore, have the following immediate corollary.

COROLLARY 3. *The class of chain graphs, \mathcal{CG} , with the LWF Markov property is stable under conditioning.*

5.2. Simultaneous marginalization and conditioning for CMGs. Corollaries 4 and 2 imply that \mathcal{CMG} with c-separation criterion is *stable under marginalization and conditioning*, which formally holds when there is a graph $H \in \mathcal{CMG}$ such that $\mathcal{J}_c(H) = \alpha(\mathcal{J}_c(G); M, C)$, where

$$\alpha(\mathcal{J}; M, C) = \{\langle A, B \mid D \rangle : \langle A, B \mid D \cup C \rangle \in \mathcal{J} \text{ and } (A \cup B \cup D) \cap (M \cup C) = \emptyset\}.$$

We now deal with the case where there are both marginalization and conditioning subsets in a CMG. We first define maximality in order to simplify the results. A graph is *maximal* if to every non-adjacent pairs of nodes, there is an independence statement associated in $\mathcal{J}(G)$. CMGs are not maximal since, for example, the class of ancestral graphs [18] is a subclass of CMGs, and there exist non-maximal ancestral graphs; see also Figure 7, for an example of a CMG that is not ancestral and that induces no independence statement of form $j \perp_c l \mid C$ for any choice of C . There is a method to generate, from a non-maximal CMG, a maximal CMG that induces the same independence model, which is beyond the scope of this manuscript. However, here we provide a sufficient condition for non-maximal graphs as a lemma, which will be used in our proofs.

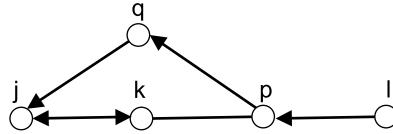


FIG. 7. A non-maximal AnG.

LEMMA 7. *If there is a collider trislide between i and j in G such that there is an arrow from an inner node of the trislide to j (or i) and $i \not\sim j$ then G is not maximal.*

We also provide the following lemma, which deals with the global behavior of the simultaneous marginalization and conditioning as described later in this section.

LEMMA 8. *There is an edge between i and j in $\alpha_{CMG}(\alpha_{CMG}(G; \emptyset, C); M, \emptyset)$ if and only if there is a walk between i and j in G on which (i) all nodes on collider sections are in $C \cup \text{ant}(C)$; (ii) on non-collider sections, (a) all nodes are in M , or (b) one endpoint is in M and also either a child of a node in M or a spouse of a node in $C \cup \text{ant}(C)$, and the other endpoint has an arrowhead at it from the adjacent node on the walk. In addition, the walk and the edge are endpoint-identical except when there is an arrowhead at the endpoint section containing i (or j), and $i \in \text{ant}(C)$ [or $j \in \text{ant}(C)$] in G .*

We now have the following important result, which illustrates that, for maximal graphs, in order to both marginalize and condition, it does not matter whether we marginalize first by using Algorithm 1 and then condition by using Algorithm 2 or vice versa.

PROPOSITION 6. *For a chain mixed graph G and two disjoint subsets M and C of its node set, it holds that*

$$\alpha_{CMG}(\alpha_{CMG}(G; M, \emptyset); \emptyset, C) = \alpha_{CMG}(\alpha_{CMG}(G; \emptyset, C); M, \emptyset)$$

if $\alpha_{CMG}(\alpha_{CMG}(G; M, \emptyset); \emptyset, C)$ is maximal.

It is also clear from the proof that if we drop the maximality assumption then the two concerned graphs in the proposition induce the same independence models. In addition, we show that the corresponding algorithm (Algorithm 1 followed by Algorithm 2 or vice versa) is well-defined for maximal graphs. We denote the corresponding function by $\alpha_{CMG}(G; M, C)$. In general, one can first apply Algorithm 2 followed by Algorithm 1, in which case we showed in the proof that an edge is present between the endpoints of the walk described in Lemma 7.

THEOREM 5. *For a chain mixed graph G and disjoint subsets M, M_1, C and C_1 of its node set,*

$$\alpha_{\text{CMG}}(\alpha_{\text{CMG}}(G; M, C); M_1, C_1) = \alpha_{\text{CMG}}(G; M \cup M_1, C \cup C_1)$$

if the two graphs are maximal.

PROOF. The result follows from the definition and Proposition 6, Theorem 3 and Theorem 1. \square

In Proposition 2, we showed that all CGs after marginalization are mapped onto \mathcal{H} , which is a subclass of CMGs. Here, we show that CGs after marginalization and conditioning are also mapped onto \mathcal{H} .

PROPOSITION 7. *The map α_{CMG} maps \mathcal{CG} and two subsets of the node set of its members surjectively onto \mathcal{H} .*

We are now ready to provide the main result, which illustrates that by applying Algorithm 1 followed by Algorithm 2 (or vice versa), we obtain the marginal and conditional independence model for a CMG (or a CG) after marginalization and conditioning.

THEOREM 6. *For a chain mixed graph G and disjoint subsets A, B, M, C and C_1 of its node set,*

$$\langle A, B | C_1 \rangle \in \mathcal{J}_c(\alpha_{\text{CMG}}(G; M, C)) \iff \langle A, B | C \cup C_1 \rangle \in \mathcal{J}_c(G).$$

PROOF. By definition and Proposition 6, Theorem 4 and Theorem 2, it is implied that

$$\begin{aligned} \langle A, B | C_1 \rangle \in \mathcal{J}_c(\alpha_{\text{CMG}}(G; M, C)) &= \mathcal{J}_c(\alpha_{\text{CMG}}(\alpha_{\text{CMG}}(G; M, \emptyset); \emptyset, C)) \\ &\iff \langle A, B | C \cup C_1 \rangle \in \mathcal{J}_c(\alpha_{\text{CMG}}(G; M, \emptyset)) \\ &\iff \langle A, B | C \cup C_1 \rangle \in \mathcal{J}_c(G). \end{aligned} \quad \square$$

6. Anterital graphs. The definition of CMGs can be considered a generalization of the definition of *summary graphs* (SGs) by [26]: CMGs collapse to SGs when there are no arrowheads pointing to lines. CMGs are also analogous to SGs in the sense that they capture the marginal and conditional models for CGs, and SGs capture the marginal and conditional models for DAGs; and CMGs exclude graphs with semi-directed cycles while SGs exclude graphs with directed cycles.

The class of ancestral graphs, defined by [18], captures the same independence models as those of SGs, but has a simpler structure than SGs. In this section, we

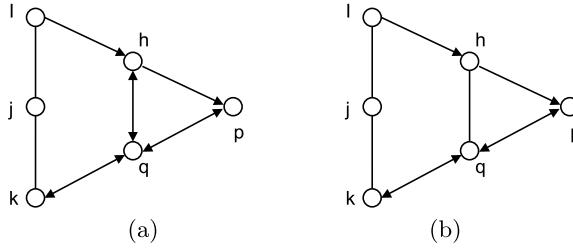


FIG. 8. (a) An AnG. (b) A CMG that is not an AnG.

define the class of *anterial graphs* (AnGs), which can be thought of as a generalization of and analogous to ancestral graphs with the same relationship to CMGs as that of ancestral graphs to SGs.

An anterial graph is a mixed graph that contains neither semi-directed cycles that contain at least an arrow; nor does it contain arcs with one endpoint that is an anterior of the other endpoint. This implies that, unlike CMGs, AnGs are simple graphs. For example, in Figure 8(a) the graph is an AnG, but in Figure 8(b) the graph is not an AnG because of the existence of the arc kq , where $k \in \text{ant}(q)$ via the semi-directed path $\langle k, j, l, h, q \rangle$ as well as the arc qp , where $q \in \text{ant}(p)$.

Here, we show that, from an anterial graph and after marginalization and conditioning, how to generate an anterial graph with the corresponding marginal and conditional independence model.

ALGORITHM 3 [$\alpha_{\text{AnG}}(G; M, C)$: Generating an AnG from an anterial graph G].

Start from G .

1. Apply Algorithm 2.
2. Apply Algorithm 1.
3. Generate respectively arrows from j to i or arcs between i and j for trislides $j \rightarrow o \dots \rightarrow i \leftrightarrow k$ or $j \leftrightarrow o \dots \rightarrow i \leftrightarrow k$ when $k \in \text{ant}(i)$ if the arrow or the arc does not already exist.
4. Generate respectively an arrow from j to i or an arc between i and j for trislides $j \rightarrow k_1 \dots \rightarrow k_m \leftrightarrow i$ or $j \leftrightarrow k_1 \dots \rightarrow k_m \leftrightarrow i$ when there is an $1 \leq r \leq m$ such that $k_r \in \text{ant}(i)$ if the arrow or the arc does not already exist. Continually apply this step until it is not possible to apply it further.
5. Remove the arc between j and i in the case that $j \in \text{ant}(i)$, and replace it with an arrow from j to i if the arrow does not already exist; and remove the arc between j and i in the case that $j \in \text{ant}(i)$ and $i \in \text{ant}(j)$, and replace it with a line between i and j if the line does not already exist.

Notice that, as we will see, steps 3, 4 and 5 of Algorithm 3 generate, from the generated CMG after step 2, an AnG that captures the same independence model

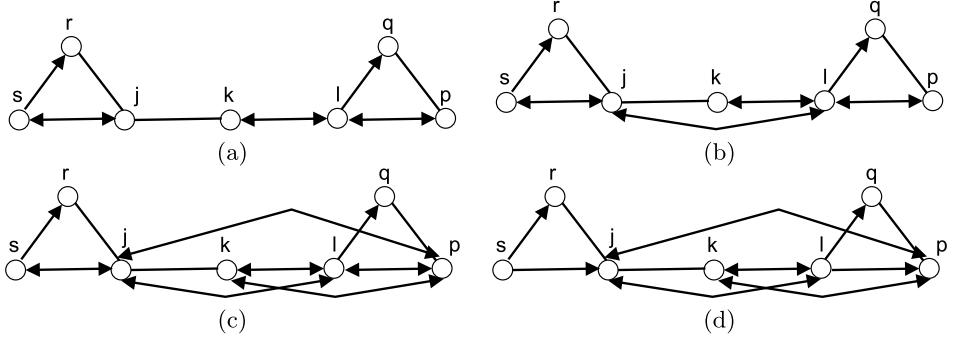


FIG. 9. (a) A chain mixed graph G . (b) The graph after applying step 3 of Algorithm 3. (c) The graph after applying step 4. (d) The generated AnG after applying step 5.

as that of the CMG. In addition, in step 4, one k_r being in $\text{ant}(i)$ implies that all k_r , $1 \leq r \leq m$, are in $\text{ant}(i)$, and in this sense we can say that a section is in $\text{ant}(i)$.

This algorithm is a generalization of the related algorithm for ancestral graphs [18, 19]. Again, one can see that sections here are treated in the same way as nodes in the ancestral-graph-generating algorithms. The idea here is that step 4 generates a direct dependency between j and i (in fact the dependency already exists) before step 5 makes the graph anterinal.

Figure 9 illustrates how to apply these steps to a CMG. We consider Algorithm 3 a function denoted by α_{AnG} . Notice that for every anterinal graph G , it holds that $\alpha_{\text{AnG}}(G; \emptyset, \emptyset) = G$. We again follow a parallel theory as that in the previous sections.

PROPOSITION 8. *Graphs generated by Algorithm 3 are AnGs.*

We first provide two lemmas that deal with the global behavior of the algorithm.

LEMMA 9. *Let H be a chain mixed graph. It holds that $i \in \text{ant}(j)$ in H if and only if $i \in \text{ant}(j)$ in the anterinal graph generated after applying steps 3, 4 and 5 of Algorithm 3 to H .*

Denote by a walk between i and j on which all sections are collider and every inner section is in $\text{ant}(i)$ a *subprimitive inducing walk* from j to i . This is a special case of a generalization of primitive inducing paths, defined in [18], where all nodes are anteriors of one of the endpoints, not either of the endpoints. We also denote the function corresponding to steps 3, 4 and 5 of Algorithm 3 by $\alpha_{\text{CMG.AnG}}$. Notice that $\alpha_{\text{AnG}}(G; M, C) = \alpha_{\text{CMG.AnG}}(\alpha_{\text{CMG}}(G; M, C))$.

LEMMA 10. *Let H be a chain mixed graph. There is an edge between i and j in $\alpha_{\text{CMG.AnG}}(H)$ if and only if there is a sub-primitive inducing walk from j to i .*

in H (which might also contain i as an inner node) with single-element endpoint sections. In addition, the edge and the walk are endpoint-identical except when $i \in \text{ant}(j)$ or $j \in \text{ant}(i)$ in H , in which case there is no arrowhead at i or at j , respectively, on the ij edge in $\alpha_{\text{CMG.AnG}}(H)$.

We now prove that Algorithm 3 does not need to be applied to an anteriel graph, but it can be applied to a chain mixed graph.

LEMMA 11. *Let H be a chain mixed graph and M and C be two subsets of its node set. It holds that $\alpha_{\text{AnG}}(\alpha_{\text{CMG.AnG}}(H); M, C) = \alpha_{\text{AnG}}(H; M, C)$.*

THEOREM 7. *For an anteriel graph G and disjoint subsets M, M_1, C and C_1 of its node set,*

$$\alpha_{\text{AnG}}(\alpha_{\text{AnG}}(G; M, C); M_1, C_1) = \alpha_{\text{AnG}}(G; M \cup M_1, C \cup C_1),$$

if the two graphs are maximal.

PROOF. Using Theorem 5 and Lemma 11, we have the following:

$$\begin{aligned} \alpha_{\text{AnG}}(\alpha_{\text{AnG}}(G; M, C); M_1, C_1) &= \alpha_{\text{AnG}}(\alpha_{\text{CMG.AnG}}(\alpha_{\text{CMG}}(G; M, C)); M_1, C_1) \\ &= \alpha_{\text{AnG}}(\alpha_{\text{CMG}}(G; M, C); M_1, C_1) \\ &= \alpha_{\text{CMG.AnG}}(\alpha_{\text{CMG}}(\alpha_{\text{CMG}}(G; M, C); M_1, C_1)) \\ &= \alpha_{\text{CMG.AnG}}(\alpha_{\text{CMG}}(G; M \cup M_1, C \cup C_1)) \\ &= \alpha_{\text{AnG}}(G; M \cup M_1, C \cup C_1). \end{aligned} \quad \square$$

Denote the set of all AnGs by \mathcal{ANG} .

PROPOSITION 9. *Let \mathcal{K} be the subset of \mathcal{ANG} with the following properties:*

1. *There is no collider trislide of form $k \longleftrightarrow i — \dots — j \leftarrow l$ unless there is an arrow from l to i .*
2. *There is no collider trislide of form $k \longleftrightarrow i — \dots — j \longleftrightarrow l$ unless there are jk and il arcs and an ij line.*

Then α_{AnG} maps \mathcal{CG} and two subsets of the node set of its members surjectively onto \mathcal{K} .

THEOREM 8. *For an anteriel graph G and disjoint subsets A, B, M, C and C_1 of its node set,*

$$\langle A, B | C_1 \rangle \in \mathcal{J}_c(\alpha_{\text{AnG}}(G; M, C)) \iff \langle A, B | C \cup C_1 \rangle \in \mathcal{J}_c(G).$$

COROLLARY 4. *The class of anteriel graphs, \mathcal{ANG} , with c-separation criterion is stable under marginalization and conditioning.*

7. Probabilistic independence models for CMGs and AnGs and comparison to other types of graphs. The most interesting independence models are induced by probability distributions. Consider a set V and a collection of random variables $(X_\alpha)_{\alpha \in V}$ with joint density f_V . By letting $X_A = (X_v)_{v \in A}$ for each subset A of V , we then use the short notation $A \perp\!\!\!\perp B | C$ for $X_A \perp\!\!\!\perp X_B | X_C$ and disjoint subsets A, B and C of V .

For a given independence model \mathcal{J} , a probability distribution P is called *faithful* with respect to \mathcal{J} if, for random vectors X_A, X_B and X_C with probability distribution P ,

$$A \perp\!\!\!\perp B | C \quad \text{if and only if} \quad \langle A, B | C \rangle \in \mathcal{J}.$$

We say that \mathcal{J} is *probabilistic* if there is a distribution P that is faithful to \mathcal{J} .

From a given collection of random variables $(X_\alpha)_{\alpha \in V}$ with a probability distribution P , one can induce an independence model $\mathcal{J}(P)$ by demanding

$$\text{if } A \perp\!\!\!\perp B | C \quad \text{then } \langle A, B | C \rangle \in \mathcal{J}(P).$$

Notice that $\mathcal{J}(P)$ is obviously probabilistic.

For a chain graph G , we say that a probability distribution with density f factorizes with respect to G if

$$f(x) = \prod_{\tau \in \mathcal{T}} f(x_\tau | x_{\text{pa}(\tau)}),$$

where \mathcal{T} is the set of chain components of G ; and

$$f(x_\tau | x_{\text{pa}(\tau)}) = \prod_a \phi_a(x),$$

where a varies over all subsets of $\tau \cup \text{pa}(\tau)$ that are complete in the moral graph of the subgraph of G induced by $\tau \cup \text{pa}(\tau)$, and $\phi_a(x)$ is a function that depends on x through x_a only; see [9] for more discussion.

Now let $\alpha(P; M, C)$ be the probability distribution obtained by usual probabilistic marginalization and conditioning for the probability distribution P . It is easy to show that if P is faithful to \mathcal{J} then $\alpha(P; M, C)$ is faithful to the marginal and conditional independence model $\alpha(\mathcal{J}; M, C)$; see Theorem 7.1 and Corollary 7.3 of [18].

It is also known that if G is a CG then there is a regular Gaussian distribution that is faithful to it. In fact, almost all the regular Gaussian distributions that factorize with respect to a CG are faithful to it; see [15]. In other words, the independence mode $\mathcal{J}_c(G)$ is probabilistic.

By Propositions 2, 7 and 9, a considerably large subclass of CMGs or AnGs are obtained by chain graphs after marginalization and conditioning. Hence, it is implied by the discussion above that for a graph H in these subclasses, $\mathcal{J}_c(H)$ is probabilistic; that is, there is a distribution (in fact at least a Gaussian distribution) that is faithful to it.

One can obtain the same result for the strictly positive discrete probability distributions since there is such a distribution that is faithful to a given CG [24]. These results motivate the use of CMGs and AnGs.

The next, and probably more important, question in order to justify the use of these classes is whether it is possible to find a parametrization, for example, Gaussian or discrete, of these graphs.

In the Gaussian case, there exists a known parametrization for the regular Gaussian distributions that factorize with respect to a CG; see [28] and [15] for two slightly different but equivalent parametrizations. For maximal ancestral graphs (MAGs), there is a known Gaussian parametrization [18]. We believe that it is possible to extend this parametrization to the class of maximal AnGs. Here is some possible actions in order to generalize this parametrization.

Notice first that the classes of CMGs and AnGs are not maximal, as explained in Section 5.2. However, as mentioned before, there is a method to generate, from non-maximal CMGs and AnGs, maximal CMGs and AnGs that induce the same independence models. Hence, one can then focus on the class of maximal AnGs.

Considering the Gaussian parametrization for MAGs, one then needs to define, instead of one matrix for the undirected part of the MAG, one symmetric matrix for every chain component of the maximal AnG (as it is done in the Gaussian parametrization for CGs). It is also needed to generalize the ordering associated to MAGs, for example, by defining an ordering for chain components containing lines instead of an ordering for the nodes. One may then follow the method described in Section 8 of the mentioned paper.

Since both parametrizations for CGs and MAGs are curved exponential families, and consequently the models associated with them are identifiable, the generalization for AnGs seems to preserve this desirable property.

Introducing a discrete parametrization for CMGs or AnGs seems much trickier. Similar to the Gaussian case, the goal should be to find a combination of discrete parametrizations for CGs (see, e.g., [14]) and summary graphs (or alternatively ADMGs—see [4]). For CMGs, a parametrization may be derived from the original CG with the use of structural equation models with latent variables. This can be considered a generalization of the method utilized in summary graph models.

Nonetheless, we again stress the importance of introducing different smooth parametrizations for CMGs and AnGs in a future work as well as studying additional non-independence constraints that arise in such models.

Besides the relevant parametrizations, it is clear that CMGs act similarly to summary graphs in the problem of marginalization and conditioning for DAGs, and AnGs act similarly to ancestral graphs. To give a more detailed comparison between CMGs (and AnGs) and summary graphs (and ancestral graphs), we first note that the lines in all these graphs have the same meaning. As mentioned before, there are no arrowheads at lines in the latter types, and one can think of sections with arrowheads pointing to them in the former types in the same manner as the

nodes in the latter types. Indeed summary graphs and ancestral graphs are subclasses of CMGs and AnGs, respectively, thus every summary or ancestral graph model is a CMG or AnG model.

In addition, in CMGs, for a collider trislide of from $i \rightarrow j \longrightarrow l \longleftarrow k$, it holds that $i \not\perp\!\!\!\perp cl$, $i \not\perp\!\!\!\perp cl \mid j$, but $i \perp\!\!\!\perp cl \mid \{j, k\}$. However, there is no summary graph that can capture the same independencies and dependencies. Hence, for any induced path with 4 nodes (and, of course, for longer paths), one can provide a CMG that is associated to a different model than summary graph models. By this, it is clear that the class of CMG models is rich in the sense that when the number of nodes grows, the number of distinct CMG models grows faster than the number of distinct summary graph models.

The class of marginal AMP chain graphs (MAMP CGs) deals with a similar problem of marginalization for AMP chain graphs. The lines in these graphs have a different meaning in independence interpretation (they are related to lines in AMP CGs), and naturally the class of models they represent is quite different. However, both classes of models contain the class of regression graph models [27], which itself contains the classes of undirected (concentration) graph models and the class of multivariate regression chain graph models as a subclass. In fact, if in a CMG, there is a section with non-adjacent endpoints that is larger than a single node then it can be seen that no MAMP CG can induce the same independence statements. This implies that, in the intersection of CMG and MAMP CG models, there is no arrowhead pointing to lines (in CMG sense). Therefore, this intersection is the same as the intersection of maximal ancestral graph and MAMP CG models (since MAMP CGs are maximal, and maximal summary and ancestral graphs induce the same independence model).

Acknowledgements. The author is grateful to Steffen Lauritzen and Thomas Richardson for helpful discussions, Nanny Wermuth for helpful discussions and comments and anonymous referees for the most helpful comments, especially detecting an error in the results.

SUPPLEMENTARY MATERIAL

Proofs (DOI: [10.1214/16-AOS1451SUPP](https://doi.org/10.1214/16-AOS1451SUPP); .pdf). We provide proofs of non-trivial lemmas, propositions and theorems in the paper as well as some more technical and yet less informative lemmas that are used in the proofs.

REFERENCES

- [1] ANDERSSON, S. A., MADIGAN, D. and PERLMAN, M. D. (2001). Alternative Markov properties for chain graphs. *Scand. J. Statist.* **28** 33–85. [MR1844349](#)
- [2] COX, D. R. and WERMUTH, N. (1993). Linear dependencies represented by chain graphs. *Statist. Sci.* **8** 204–218, 247–283. With comments and a rejoinder by the authors. [MR1243593](#)

- [3] DRTON, M. (2009). Discrete chain graph models. *Bernoulli* **15** 736–753. [MR2555197](#)
- [4] EVANS, R. J. and RICHARDSON, T. S. (2014). Markovian acyclic directed mixed graphs for discrete data. *Ann. Statist.* **42** 1452–1482. [MR3262457](#)
- [5] FRYDENBERG, M. (1990). The chain graph Markov property. *Scand. J. Statist.* **17** 333–353. [MR1096723](#)
- [6] GEIGER, D., HECKERMAN, D., KING, H. and MEEK, C. (2001). Stratified exponential families: Graphical models and model selection. *Ann. Statist.* **29** 505–529. [MR1863967](#)
- [7] KIVIARI, H., SPEED, T. P. and CARLIN, J. B. (1984). Recursive causal models. *J. Austral. Math. Soc. Ser. A* **36** 30–52. [MR0719999](#)
- [8] KOSTER, J. T. A. (2002). Marginalizing and conditioning in graphical models. *Bernoulli* **8** 817–840. [MR1963663](#)
- [9] LAURITZEN, S. L. (1996). *Graphical Models*. Oxford Statistical Science Series **17**. The Clarendon Press, Oxford. [MR1419991](#)
- [10] LAURITZEN, S. L. and SPIEGELHALTER, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *J. Roy. Statist. Soc. Ser. B* **50** 157–224. [MR0964177](#)
- [11] LAURITZEN, S. L. and WERMUTH, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Statist.* **17** 31–57. [MR0981437](#)
- [12] MARCHETTI, G. M. and LUZZARELLI, M. (2011). Chain graph models of multivariate regression type for categorical data. *Bernoulli* **17** 827–844. [MR2817607](#)
- [13] PEARL, J. (2009). *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge Univ. Press, New York. [MR2548166](#)
- [14] PEÑA, J. M. (2009). Faithfulness in chain graphs: The discrete case. *Internat. J. Approx. Reason.* **50** 1306–1313. [MR2556122](#)
- [15] PEÑA, J. M. (2011). Faithfulness in chain graphs: The Gaussian case. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS 2011)* **15** 588–599. [JMLR.org](#).
- [16] PEÑA, J. M. (2014). Marginal AMP chain graphs. *Internat. J. Approx. Reason.* **55** 1185–1206. [MR3197803](#)
- [17] RICHARDSON, T. (2003). Markov properties for acyclic directed mixed graphs. *Scand. J. Stat.* **30** 145–157. [MR1963898](#)
- [18] RICHARDSON, T. and SPIRES, P. (2002). Ancestral graph Markov models. *Ann. Statist.* **30** 962–1030. [MR1926166](#)
- [19] SADEGHI, K. (2013). Stable mixed graphs. *Bernoulli* **19** 2330–2358. [MR3160556](#)
- [20] SADEGHI, K. (2016). Supplement to “Marginalization and conditioning for LWF chain graphs”. DOI:10.1214/16-AOS1451SUPP.
- [21] SHPITSER, I. and PEARL, J. (2008). Dormant independence. In *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence* 1081–1087. AAAI Press, Menlo Park.
- [22] STUDENÝ, M. (1998). Bayesian networks from the point of view of chain graphs. In *UAI* 496–503. Morgan Kaufmann, San Francisco, CA.
- [23] STUDENÝ, M. (2005). *Probabilistic Conditional Independence Structures*. Springer, London. [MR3183760](#)
- [24] STUDENÝ, M. and BOUCKAERT, R. R. (1998). On chain graph models for description of conditional independence structures. *Ann. Statist.* **26** 1434–1495. [MR1647685](#)
- [25] VERMA, T. and PEARL, J. (1990). Equivalence and synthesis of causal models. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence (UAI-90)* 220–227. Elsevier, New York.
- [26] WERMUTH, N. (2011). Probability distributions with summary graph structure. *Bernoulli* **17** 845–879. [MR2817608](#)

- [27] WERMUTH, N. and SADEGHI, K. (2012). Sequences of regressions and their independences. *TEST* **21** 215–252. [MR2935353](#)
- [28] WERMUTH, N., WIEDENBECK, M. and COX, D. R. (2006). Partial inversion for linear systems and partial closure of independence graphs. *BIT* **46** 883–901. [MR2285213](#)

STATISTICAL LABORATORY
CENTRE FOR MATHEMATICAL STUDIES
WILBERFORCE ROAD
CAMBRIDGE, CB3 0WB
UNITED KINGDOM
E-MAIL: k.sadeghi@statslab.cam.ac.uk