Independence, Decomposability and functions which take values into an Abelian Group

Adrian Silvescu
Department of Computer Science
Iowa State University
Ames, IA 50010, USA
silvescu@cs.iastate.edu

Vasant Honavar
Department of Computer Science
Iowa State University
Ames, IA 50010, USA
honavar@cs.iastate.edu

Abstract

Decomposition is an important property that we exploit in order to render problems more tractable. The decomposability of a problem implies the existence of some "independences" between the variables that are in play in the problem under consideration. In this paper we investigate the decomposability of functions which take values into an Abelian Group. For the problems which involve these types of functions we define a notion of conditional independence between subsets of the problem's variables. We then prove a decomposition theorem that relates independences and a factorisation property of functions taking values into an Abelian Group. As particular cases of this theorem we retrieve the Hammersly-Clifford theorem for probability distributions, an Additive Decomposition theorem for functions such as value functions or fitnesses and a relational algebra decomposition theorem.

1 Introduction[1]

Probabilistic Graphical Models (PGMs) have proved to be a successful way of representing probability distributions in a concise and intuitive form. Compact graphical representations support efficient reasoning and learning algorithms in many cases that arise in practice [Pearl '88, Cowell et al '99, Jordan '05]. The key trick behind PGMs is the notion of probabilistic independence. By the means of independence we can "decompose" the probability distribution into smaller parts, and thus considerably reduce dimensionality, in terms of the number of independent parameters that we need to know in order to specify the probability distribution.

[1]Two more pages with supplementary information at the discretion of the reviewers are provided in the Appendix. Please note that without these pages the paper length is exactly 8 pages.

These desirable properties of PGMs make us wonder: Can we have a similar story for other or more general cases? [Bacchus and Grove '95] examined conditionally additive decomposable representations for utility functions and proved results similar to those that hold in PGMs . Others have also have done work in this area. See [Jensen '01] for starters, and also the references in [Bacchus and Grove '95] for prior work in this vein.

In this paper, we will examine the decomposability of functions which take values into an Abelian group. Let us call this class of functions $\mathcal{F}_{\cdot\to AG}$. $\mathcal{F}_{\cdot\to AG}$ includes as particular cases: strictly positive probability distributions; additive decomposable functions and relations among others. We will show that for $\mathcal{F}_{\cdot\to AG}$ we can define a general notion of Conditional Independence that is the natural generalization of probabilistic Conditional Independence. This Conditional Independence will allow us to *"decompose"* a function f from $\mathcal{F}_{\cdot\to AG}$ in the same way as a probability distribution can be "decomposed". More exactly, we prove a generalization of the well known Hammersly-Clifford theorem, which will hold for arbitrary functions from $\mathcal{F}_{\cdot\to AG}$.

Aside from the general interest, these results will facilitate for techniques mainly developed in the PGMs framework to be emulated in any of the areas where the problem under consideration can be cast as a function which takes values into an Abelian Group.

The reminder of the paper is organized as follows: In the next section we will outline some definitions regarding Decomposition and Independence in very general terms. Then in section 3 we will shrink our domain of interest to functions whose ranges are Abelian Groups: $\mathcal{F}_{\cdot\to AG}$. In such a setup we will define a general Conditional Independence concept with respect to a function f and explore some of the properties of this new concept. In the same section we will also prove the main result of this paper which is a factorisation property consisting in the natural generalization of the Hammersly-Clifford theorem [Besag '74] for arbitrary functions from $\mathcal{F}_{\cdot\to AG}$. Then is section 4 we present some interesting particular cases of decomposable functions which take values into an Abelian Group, such as: probability distributions, additive decomposable functions and relations. Finally in section 5 we conclude with a discussion and potential further work.

2 Decomposition and Independence

Decomposition is an important property that renders problems more tractable. Decomposition exploits the fact that occasionally a problem can be split (*decomposed*) into two parts which can be solved in isolation and then the overall solution can be obtained by aggregating the partial solutions. If such is the case, we say that the two parts are *independent* with respect to the problem under consideration.

The two parts, in practice, are seldom disjoint, but all is not lost in this case, because we can define a weaker notion of independence, namely conditional independence, that is still useful.

Furthermore, if one or both of the parts are in turn decomposable into smaller parts, we can apply the strategy of aggregating partial solutions, recursively. Thus, in a divide and conquer fashion, we would be able to obtain a solution for our problem by aggregating it from solutions of smaller and smaller parts.

These are the general intuitions underlying Decomposition and Independence. Next we will try to capture these intuitions with more precise definitions.

Definition (problem): A problem P is a triple $P = (D, S, sol_P)$ where D is a set called the *domain set*, S is a set called the *solutions set* and $sol_P : D \to S$ is a function that maps an element d from the domain of the problem to its solution s.

Example: $Determinant\_Computation(\mathcal{M}^2, \mathbb{R}, det)$ where $\mathcal{M}^2$ is the set of all square matrices, and det is the function that returns the determinant of a matrix.

Definition (conditional independence - value-based) Given a problem $P = (D, S, sol_P)$ we say that an object $d \in D$ is decomposable into a and b conditioned on c or equivalently, a is independent of b conditioned on c, both with respect to the problem P, if there exist two problems $P_1 = (A \times C, S_{AC}, sol_{P_1})$ and $P_2 = (B \times C, S_{BC}, sol_{P_2})$ and an operator $\oplus_P^{P_1,P_2} : S_{AC} \times S_{BC} \to S_D$ such that $(a, c) \in A \times C$, $(b, c) \in B \times C$ and:

$$sol_P(d) = sol_{P_1}(a, c) \oplus_P^{P_1,P_2} sol_{P_2}(b, c)$$

■

We will call this notion of independence *value-based independence* (a.k.a. context-specific independence [Boutilier et al '96] in the probabilistic setting).

A stronger notion of independence / decomposition is the *variable-based independence*:

Definition (conditional independence - variable-based) Given a problem $P = (D, S, sol_P)$ we say that, D is decomposable into A and B conditioned on C, or equivalently, that A is independent of B conditioned on C , both with respect to the problem P, if $D = A \times B \times C$ (or more liberally D is isomorph with $A \times B \times C$) and there exist two problems $P_1 = (A \times C, S_{AC}, sol_{P_1})$ and $P_2 = (B \times C, S_{BC}, sol_{P_2})$ and an operator $\oplus_P^{P_1,P_2} : S_{AC} \times S_{BC} \to S_D$ such that for all $d = (a, b, c) \in A \times B \times C$ we have:

$$sol_P(d = (a, b, c)) = sol_{P_1}(a, c) \oplus_P^{P_1,P_2} sol_{P_2}(b, c)$$

where $\oplus_P^{P_1,P_2} : S_{AC} \times S_{BC} \to S_D$. ■

In the rest of this paper we will be concerned with variable-based independence. More exactly, we will consider the particular case when the operator $\oplus_P^{P_1,P_2}$ is an operator $\oplus_P^{P_1,P_2} : G \times G \to G$. That is, the solutions to the problems P, P_1, P_2 are elements from the same set G . Furthermore, we will assume that the operator $\oplus_P^{P_1,P_2}$ does not depend on the problems P, P_1, P_2 and therefore we can drop them from the notation since we will have only one operator regardless: $\oplus : G \times G \to G$. Lastly, we will consider G to be an Abelian Group for reasons that will become apparent lately.

3 Independence & Decomposition in $\mathcal{F}_{\cdot \to AG}$

In this section we consider Independence and Decomposability pertaining to functions whose ranges are Abelian Groups: $\mathcal{F}_{\cdot \to AG}$. We start with the definition and examples of Abelian Groups (3.1). We then define a general notion of Conditional Independence with respect to a function f, $f \in \mathcal{F}_{\cdot \to AG}$, denoted $I_f(\cdot, \cdot|\cdot)$, and show how it subsumes probabilistic independence, additive independence and relational independence (3.2). Then we prove that $I_f(\cdot, \cdot|\cdot)$ satisfies the following four properties: trivial independence, symmetry, weak union and intersection, which are cherished as essential/defining properties for the notion of independence [Pearl '88] (3.3). This is the place where the Abelian Group properties of G will come into play. Capitalizing on the four independence properties we move along and establish the equivalence of Global, Local and Pairwise Markov Properties for our $I_f(\cdot, \cdot|\cdot)$, which follows as a consequence of these properties [Pearl and Paz '87] (3.4). Markov Properties in place, we move on and prove the main theorem of this paper (3.5). This theorem allows us to factorize the function f over the maximal cliques of an associated Markov Network graph which reflects the Conditional Independences satisfied by the function f . This theorem is the natural generalisation of the Hammersly-Clifford theorem from probabilities to the more general case of $\mathcal{F}_{\cdot \to AG}$.

3.1 Abelian Groups

In this section we give the definition for Abelian Groups, followed by some illustrative examples.

Definition. **(Abelian Group)** An Abelian Group is a quadruple $(G, \oplus, \theta, \ominus)$ where G is a nonempty set, $\oplus : G \times G \to G$ is a binary operation over elements from G that returns an element of G, $\theta \in G$ and $\ominus : G \to G$ is a unary operation over the elements of G that returns an element of G, with the following properties:

1. $\oplus$ is associative i.e., $\forall g_1, g_2, g_3 \in G\ (g_1 \oplus (g_2 \oplus g_3)) = ((g_1 \oplus g_2) \oplus g_3)$
2. $\oplus$ is commutative i.e., $\forall g_1, g_2 \in G\ g_1 \oplus g_2 = g_2 \oplus g_1$
3. θ is an identity element i.e., $\forall g \in G\ g \oplus \theta = \theta \oplus g = g$
4. $\ominus$ is an inversion operator i.e., $\forall g \in G\ \exists! h \in G\ g \oplus h = h \oplus g = \theta$. We will call $\ominus g$ the unique h with the previous property. Subsequently, the inversion property can be written as $\forall g \in G\ \exists! \ominus g \in G\ g \oplus \ominus g = \ominus g \oplus g = \theta$

Examples: 1. $(\mathbb{R}, +, 0, -)$ is a group, where: $\mathbb{R}$ is the set of real numbers; $+$ is the addition between real numbers; 0 is zero; and $-$is the unary operator minus that returns the inverse of a real number (e.g., $-(7) = -7$ and $-(-6) = 6$). In more common notation we use $-$ as a binary operator where $a - b$ actually stands for $a + -b$.

2. $((0, \infty], \cdot, 1, \ ^{-1})$ is a group. Where: $\cdot$ is the multiplication between real numbers; 1 is one; and $\ ^{-1}$is the inverse of a real number with respect to multiplication (i.e, $a^{-1} = \frac{1}{a}$). In more common notation we use fractions as binary operators therefore having expressions such as $\frac{a}{b}$, which stands for $a \cdot b^{-1}$.

3. $(\mathbb{R}, \cdot, 1, \ ^{-1})$ is not a group. This is because 0 has no inverse.

4. $\mathbb{Z}_2 = (\{0, 1\}, \otimes, 0, -)$ is a group where $\otimes$ stands for the Exclusive OR (XOR) operation (or equivalently, addition modulo 2). More exactly: $0 \otimes 0 = 0$; $0 \otimes 1 = 1$;$1 \otimes 0 = 1$ and $1 \otimes 1 = 0$, and $-$ is the identity operator, that is: $-0 = 0$; and $-1 = 1$. ■

For the purposes of simplifying notation, for the rest of this section, we will use as operators the standard operations of the Additive Abelian Group instead of the fancier ones that we have introduced in the definition of a group. More precisely, instead of saying: let $(G, \oplus, \theta, \ominus)$ be a group ..., we will say: let $(G, +, 0, -)$ be a group ... . This will make the definitions and proofs look more familiar since they are in additive notation. However the only properties that we will use are those of groups and as a consequence all the results will hold for arbitrary groups, such as, for example, the multiplicative or the $\mathbb{Z}_2$ group. Additionally, we will also use the shortcut notation of $a \ominus b$ to stand for $a \oplus \ominus b$, which in our familiar additive notation, to be used from now on, will be nothing but $a - b$ (which stands for $a + -b$).

3.2 Independence with respect to a function f

In this subsection we will define a general notion of Conditional Independence with respect to a function f, $I_f(\cdot, \cdot|\cdot)$. We start with some notations, then present the definition and finish with some illustrative examples.

Notation. Let $(X_\alpha)_{\alpha \in V}$ stand for a collection of variables that take values into the spaces $(\mathcal{X}_\alpha)_{\alpha \in V}$, where V is a set of indices for these variables. For a subset A of V let $\mathcal{X}_A = \times_{\alpha \in A} \mathcal{X}_\alpha$ and in particular let $\mathcal{X}$ stand for $\mathcal{X}_V$. The collection $(X_\alpha)_{\alpha \in V}$ represents the relevant variables pertaining to our problem. $\mathcal{X}_A$ will stand for the set of all possible configurations for the variables indexed by A. Typical elements of $\mathcal{X}_A$ will be denoted as $x_A = (x_\alpha)_{\alpha \in A}$. Similarly, X_A will stand for $(X_\alpha)_{\alpha \in A}$ and X will stand for X_V. Given $A, B, C \subseteq V$ some sets of variable indices we will make conditional independence statements regarding the associated subsets of variables X_A,X_B,X_C such as: conditioned on X_C, the sets of variables X_A and X_B are independent and write $I(X_A, X_B|X_C)$. We will actually use the shorthand formulation, conditioned on C, A and B are independent and the shorthand notation $I(A, B|C)$ to stand for $I(X_A, X_B|X_C)$.

Definition (Conditional Independence with respect to a function f -$I_f(\cdot, \cdot|\cdot)$): Let $(G, +, 0, -)$ be a group, $(X_\alpha)_{\alpha \in V}$ be a collection of variables indexed by V and $\mathcal{X} = \times_{\alpha \in V} \mathcal{X}_\alpha$ be the set of configurations for these variables. Let $f : \mathcal{X} \to G$ be a function from the set $\mathcal{X}$ to G. Furthermore, let $A, B, C \subseteq V$ a partition of V (hence $\mathcal{X} = \mathcal{X}_A \times \mathcal{X}_B \times \mathcal{X}_C$). Then we say that A is independent of B conditioned on C with respect to the function f, and write $I_f(A, B|C)$, if there exist two functions f_1, f_2 such that:

$$f(X) = f(X_A, X_B, X_C) = f_1(X_A, X_C) + f_2(X_B, X_C)$$

where $f_1 : \mathcal{X}_A \times \mathcal{X}_C \to G$ and $f_2 : \mathcal{X}_B \times \mathcal{X}_C \to G$.

Instead of the previous formula we will use the shorthand notation

$$f(V) = f(A, B, C) = f_1(A, C) + f_2(B, C)$$

■

Note that in our notion of conditional independence just defined, A, B, C is necessarily a partition of V as opposed to the case of probabilistic independence where it is possible that A, B, C do not cover V . In our (general) case, if

A, B, C do not cover V then $f(A, B, C)$ is not necessarily defined. In the case of probability distribution there is a natural way to define $f(A)$ when $A \subsetneq V$ based on $f(V)$. That is, by the means of marginals. In the more general cases that we will study, (e.g., additive independence) the equivalent notion of a marginal is not necessarily present. As a consequence the theory developed in this context will be weaker, and hence more general. In the terminology of [Geiger & Pearl '93] independence statements $I(A, B|C)$ such that A, B, C cover V, are called saturated independence statements.

We proceed next with some examples of Conditional Independence with respect to a function f , $I_f(\cdot, \cdot|\cdot)$:

Probabilistic $I_f(\cdot, \cdot|\cdot)$: In the case when the group is $((0, \infty), \cdot, 1, \ ^{-1})$ and furthermore the function $f : \mathcal{X} \to (0, \infty)$ is a probability distribution (that is, $\sum_{x \in X} f(x) = 1$, or more generally $\int_X df = \int_X f(x)dx = 1$) then our notion of conditional independence $I_f(\cdot, \cdot|\cdot)$ becomes probabilistic conditional independence. Note that this group includes strictly positive probabilities only, in order to satisfy the group property (0 has no inverse element). $I_f(A, B|C)$ in this case is equivalent with:

$$f(A, B, C) = f_1(A, C) \cdot f_2(B, C)$$

obtained by substituting in the definition the original $+$ with the probabilistic group binary operator $\cdot$. This is an alternative definition for probabilistic conditional independence. For a proof see [Lauritzen '89]. Thus, we have just shown that probabilistic conditional independence is a particular case of Conditional Independence with respect to a function f , when f happens to be a probability distribution.

Additive $I_f(\cdot, \cdot|\cdot)$: In the case the group is $(\mathbb{R}, +, 0, -)$ and we have a function $f : \mathcal{X} \to \mathbb{R}$ we obtain the notion of Additive Independence i.e.,:

$$f(A, B, C) = f_1(A, C) + f_2(B, C)$$

Relational $I_f(\cdot, \cdot|\cdot)$: In the case the group is $\mathbb{Z}_2 = (\{0, 1\}, \otimes, 0, -)$ and we have a function $f : \mathcal{X} \to \mathbb{Z}_2$ (a.k.a. relation) we obtain:

$$f(A, B, C) = f_1(A, C) \otimes f_2(B, C)$$

3.3 Properties of $I_f(\cdot, \cdot|\cdot)$

In this section we prove some properties associated with $I_f(\cdot, \cdot|\cdot)$. These are general properties that researchers [Pearl and Paz '87, Pearl '88 Geiger and Pearl '93] have identified as desirable for any conditional independence relation since they capture some intuitive notions that pertain to independence. We will show that the our Conditional Independence relation with respect a function f - $I_f(\cdot, \cdot|\cdot)$ satisfies these properties.

Theorem (independence properties): *Let $(G, +, 0, -)$ be a group, $(X_\alpha)_{\alpha \in V}$ be a collection of variables indexed by V, $f : \mathcal{X} \to G$ be a function from the set $\mathcal{X}$ to the group G, and A, B, C, D be subsets of V. Then the Conditional Independence relation with respect to f, $I_f(\cdot, \cdot|\cdot)$ has the following properties:*

1. **(Trivial Independence)** $I_f(A, \emptyset|B)$ - $\forall A, B$ *a partition of V.*
2. **(Symmetry)** $I_f(A, B|C)$ iff $I_f(B, A|C)$ - $\forall A, B, C$ *a partition of V.*
3. **(Weak Union)** $I_f(A, B \cup D|C) \Rightarrow I_f(A, B|C \cup D)$ - $\forall A, B, C, D$ *a partition of V.*
4. **(Intersection)** $I_f(A, B|C \cup D)$ & $I_f(D, B|C \cup A) \Rightarrow I_f(A \cup D, B|C)$ - $\forall A, B, C, D$ *a partition of V.*

Proof.

1. (Trivial Independence) To show that $I_f(A, \emptyset|B)$ we have to prove that there exist f_1, f_2 such that $f(A, B) = f_1(A, B) + f_2(B)$. By taking $f_1(A, B) = f(A, B)$ and $f_2(B) \equiv 0$ (0 is the identity element of G) we have the desired conclusion.

2. (Symmetry) is obvious from the definition of $I_f(A, B|C)$ and the commutativity of $+$.

3. (Weak Union)

We want to prove that: $I_f(A, B \cup D|C) \Rightarrow I_f(A, B|C \cup D)$.

$I_f(A, B \cup D|C)$ implies that $f(A, B, C, D) = f_1(A, C) + f_2(B \cup D, C) = f_1(A, C) + f_2(B, C, D)$ where the last equality follows from the fact that $B \cap D = \emptyset$.

To show that $I_f(A, B|C \cup D)$ we need to show that there exists f_3, f_4 such that $f(A, B, C, D) = f_3(A, C \cup D) + f_4(B, C \cup D) = f_3(A, C, D) + f_4(B, C, D)$ where the last equality follows from the fact that $C \cap D = \emptyset$.

By taking $f_3(A, C, D) = f_1(A, C)$ and $f_4(B, C, D) = f_2(B, C, D)$ we have that $f(A, B, C, D) = f_1(A, C) + f_2(B, C, D) = f_3(A, C, D) + f_4(B, C, D)$ which by definition implies that $I_f(A, B|C \cup D)$.

4. (Intersection)

We want to prove that $I_f(A, B|C \cup D)$ & $I_f(D, B|C \cup A) \Rightarrow I_f(A \cup D, B|C)$.

$I_f(A, B|C \cup D)$ implies that $f(A, B, C, D) = f_1(A, C \cup D) + f_2(B, C \cup D) = f_1(A, C, D) + f_2(B, C, D)$ (eq. 1)

$I_f(D, B|C \cup A)$ implies that $f(A, B, C, D) = f_3(D, C \cup A) + f_4(B, C \cup A) = f_3(A, C, D) + f_4(A, B, C)$ (eq. 2). By subtracting the two equations we get

$f(A, B, C, D) - f(A, B, C, D) = 0 = f_1(A, C, D) + f_2(B, C, D) - f_3(A, C, D) - f_4(A, B, C)$.

Let $f_5(A,C,D) = f_1(A,C,D) - f_3(A,C,D)$ and by moving $f_4(A,B,C)$ to the left hand side of the previous formula we get

$f_4(A,B,C) = f_2(B,C,D) + f_5(A,C,D)$. But D does not appears in the left hand side (f_4 does not depend on D). Let $d \in D$ an arbitrary, but fixed, element of D. Then

$f_4(A,B,C) = f_6(B,C,d) + f_7(A,C,d) = f_8(B,C) + f_9(A,C)$. Now by re-substituting $f_4(A,B,C)$ in the equation (eq. 2) we have

$f(A,B,C,D) = f_3(A,C,D) + f_8(B,C) + f_9(A,C)$. Finally by taking $f_{10}(A,C,D) = f_7(A,C) + f_3(A,C,D)$ we have

$f(A,B,C,D) = f_6(B,C) + f_{10}(A,C,D)$ which by definition implies that $I_f(A \cup D, B|C)$. ■

Note that in order to prove Trivial Independence we need the identity element property of the Abelian group $(G,+,0,-)$, to prove Symmetry we needed commutativity, and to prove Intersection we needed associativity and most importantly the inverse operator $-$. So it seems that we "need" all the Abelian Group properties.

3.4 Markovian Properties of Independence

We will now study relations between graph properties and conditional independence statements. We first start with some graph terminology, then we define different types of Markov properties and subsequently, establish their equivalence for any Independence relation satisfying the four properties introduced in the previous section.

Graph notions: A *graph* is a pair $\mathcal{G} = (V,E)$ where V is a finite set of vertices and E is a set of edges. That is, E is set of pairs of vertices $E \subseteq V \times V$. A graph is called *undirected* if it has the property that for every $\alpha, \beta \in V$ $(\alpha,\beta) \in E$ if and only if $(\beta,\alpha) \in E$. Thus for the case of undirected graphs there is no distinction between the edges (α,β) and (β,α) and we will use them interchangeably to mean the same thing, namely an undirected edge between α and β. In what follows we will only consider undirected graphs.

A graph $\mathcal{G} = (V,E)$ is called complete iff there is an edge between all of its vertices. A *subgraph* of a graph $\mathcal{G} = (V,E)$ associated with set of vertices V', $V' \subseteq V$, is a graph $\mathcal{G}' = (V',E')$ such that $E' = E \cap (V' \times V')$. A set of vertices $C \subseteq V$ is called a *clique* in the graph $\mathcal{G} = (V,E)$ if the subgraph of $\mathcal{G}$ associated with C is a complete graph. That is, there is an edge between every two vertices in C in the graph $\mathcal{G}$. A clique C is called a *maximal clique* of $\mathcal{G}$ if there is no other clique C' in the graph $\mathcal{G}$ such that $C \subset C'$. Given a graph $\mathcal{G} = (V,E)$ we will use $MaxCliques(\mathcal{G})$ to denote the set of maximal cliques of $\mathcal{G}$.

Given a set $A \subseteq V$ we denote by $\mathcal{N}(A)$ and call *neighbourhood of* A the set of vertices from $V \backslash A$ that share at least one edge with an element in A. More precisely, $\mathcal{N}(A) = \{\beta | \beta \notin A \text{ and } \exists \alpha \in A \text{ such that } (\alpha,\beta) \in E\}$.

Given two vertices $\alpha, \beta \in V$ we say that there exists a *path* between α and β if there exists a set of vertices $\gamma_1, ..., \gamma_k$, $k \geq 0$ such that $(\alpha,\gamma_1), (\gamma_i,\gamma_{i+1}), (\gamma_k,\beta) \in E$ $\forall 1 \leq i < k$. We will call the sequence $\alpha, \gamma_1, ..., \gamma_k, \beta$ the path from α to β. Furthermore, given three subsets of vertices $A, B, C \subseteq V$ we say that C *separates* A *from* B *in the graph* $\mathcal{G} = (V,E)$ if there is no path between a vertex in A to a vertex in B that does not contain vertices from C. We will use $Sep_{\mathcal{G}}(A,B|C)$ to denote the fact that C separates A from B in the graph $\mathcal{G} = (V,E)$.

Now we have all the graph concepts needed to explore the connections between conditional independence statements and separability in the associated graph.

Definition (Markov properties): [Pearl '88, Lauritzen '96] Let $\mathcal{G} = (V,E)$ be an undirected graph where V is a set of indices into a collection of variables $(X_\alpha)_{\alpha \in V}$. Then we say that the conditional independence relation has the following properties relative to the graph $\mathcal{G}$ iff:

1. (P) *Pairwise Markov Property* relative to $\mathcal{G}$ iff $(\alpha,\beta) \notin E \Rightarrow I(\alpha,\beta | V \backslash \{\alpha,\beta\})$.
2. (L) *Local Markov Property* relative to $\mathcal{G}$ iff $\forall \alpha \in V$ $I(\alpha, V \backslash (\{\alpha\} \cup \mathcal{N}(\alpha)) | \mathcal{N}(\alpha))$.
3. (G) *Global Markov Property* relative to $\mathcal{G}$ iff for any two sets $A, B \subseteq V$ such that $V \backslash (A \cup B)$ separates A and B in the graph $\mathcal{G}$ we have $I(A,B | V \backslash (A \cup B))$.

It turns out that the previous three relations are equivalent for any independence relation satisfying properties 1-4 of the previous section (trivial independence, symmetry, weak union and intersection).

Theorem (Markov properties equivalence): [Pearl and Paz '87] (G) $\Leftrightarrow$ (L) $\Leftrightarrow$ (P) *for any conditional independence relation* $I(\cdot,\cdot|\cdot)$ *that satisfies* ***Trivial independence, Symmetry, Weak union*** *and* ***Intersection***.

Proof. This theorem has been proved by [Pearl and Paz '87] and can also be found in [Pearl '88, Lauritzen '96, Jordan '05]. Note that the Global Markov property is slightly weaker in our case because we have only saturated independence and hence we cannot pick arbitrary sets that separate A and B instead of just $X \backslash (A \cup B)$. Nevertheless the equivalence holds. See Appendix / Longer version of this paper [Silvescu & Honavar '05] for a proof. ■

Corollary. (G) $\Leftrightarrow$ (L) $\Leftrightarrow$ (P) *for* $I_f(\cdot,\cdot|\cdot)$.

3.5 The factorization theorem

We will now prove a theorem that ties the Conditional Independence relation with respect to a function f, $I_f(\cdot,\cdot|\cdot)$

with a factorization property of the function f over the maximal cliques of the Pairwise Markov Property (P) associated graph, from the previous section . Note that we can use any Markov Property's associated graph since as we have just saw that they are all the same graph.

Definition (factorization property): Let $\mathcal{G} = (V, E)$ be an undirected graph, let $(G, +, 0, -)$ be a group, $(X_\alpha)_{\alpha \in V}$ be a collection of variables indexed by V and $f : \mathcal{X} \to G$ be a function from the set $\mathcal{X}$ to the group G. We say that f satisfies the factorization property (F) with respect to the graph $\mathcal{G}$ iff

$$(F) \qquad f(V) = \sum_{C \in MaxCliques(\mathcal{G})} f_C(C)$$

Theorem (factorization): *Let $\mathcal{G} = (V, E)$ be an undirected graph, $(G, +, 0, -)$ be a group, $(X_\alpha)_{\alpha \in V}$ be a collection of variables indexed by V, $f : \mathcal{X} \to G$ be a function from the set $\mathcal{X}$ to the group G. Then* (G) $\Leftrightarrow$ (L) $\Leftrightarrow$ (P) $\Leftrightarrow$ (F).

Proof. We will prove (F) $\Rightarrow$ (G) and (P) $\Rightarrow$ (F) and this will be enough to prove the theorem because the other equivalences follow form the ***Markov properties*** theorem in the previous section.

(F) $\Rightarrow$ (G) Let $\mathcal{G} = (V, E)$ be a graph and $f : \mathcal{X} \to G$ be a function that satisfies the factorization property with respect to $\mathcal{G}$. Then it follows that:

$$f(V) = \sum_{C \in MaxCliques(\mathcal{G})} f_C(C)$$

Now let A, B be two sets such that $V \backslash (A \cup B)$ separates A and B in the graph $\mathcal{G}$. Then

$$\begin{aligned} f(V) &= \sum_{C \in MaxCliques(\mathcal{G}) \,\&\, C \cap A \neq \emptyset} f_C(C) \\ &\quad + \sum_{C \in MaxCliques(\mathcal{G}) \,\&\, C \cap A = \emptyset} f_C(C) \end{aligned}$$

Let $f_1(V \backslash B) = \sum_{C \in MaxCliques(\mathcal{G}) \,\&\, C \cap A \neq \emptyset} f_C(C)$ and

$f_2(V \backslash A) = \sum_{C \in MaxCliques(\mathcal{G}) \,\&\, C \cap A \neq \emptyset} f_C(C)$. To show that f_1 and f_2 are well defined we have to show that the right hand sides of their definitions contain only variables from $V \backslash B$ and $V \backslash A$ respectively. Obviously f_2 contains only variables that are not from A. We will show that f_1 has variables from $V \backslash B$ only, by contradiction.

Suppose f_1 contains variables from B. Then it follows that there exists a clique C such that $C \in MaxCliques(\mathcal{G})$, $C \cap A \neq \emptyset$ and also $C \cap B \neq \emptyset$. Let $\alpha \in C \cap A$ and $\beta \in C \cap B$. But since C is a clique in $\mathcal{G}$ it follows that (α, β) is an edge in $\mathcal{G}$, which contradicts the fact that $V \backslash (A \cup B)$ separates A and B in the graph $\mathcal{G}$.

Now given that f_1 and f_2 are well defined we can write:

$$\begin{aligned} f(V) &= f_1(V \backslash A) + f_2(V \backslash B) \\ &= f_1(A, V \backslash (A \cup B)) + f_2(B, V \backslash (A \cup B)) \end{aligned}$$

Which implies, by definition, that $I_f(A, B | V \backslash (A \cup B))$.

(P) $\Rightarrow$ (F) In order to prove this implication we will use the following helpful lemma:

Lemma (Moebius inversion): *Let f and g be two functions defined on the set of all subsets of a finite set V of variable indices, taking values into an Abelian Group $(G, +, 0, -)$. Then the following two statements are equivalent:*

(1) *for all* $A \subseteq V : g(A) = \sum_{B : B \subseteq A} f(B)$

(2) *for all* $A \subseteq V : f(A) = \sum_{B : B \subseteq A} (-1)^{|A \backslash B|} g(B)$

where, by $(-1)^k$ we mean $-$ if k is odd and $+$ if k is even. (Note that we need this explicitation because multiplication is not necessarily defined over the elements of G)

Proof. A proof of this lemma can be found in [Lauritzen '96, Jordan '05]. See also Appendix / Longer version of this paper [Silvescu & Honavar '05]. ■

We are now ready to prove the (P)$\Rightarrow$ (F) implication from the **factorization** theorem.

Let $f : \mathcal{X} \to G$ be the function, which induces an Independence relation $I_f(\cdot, \cdot | \cdot)$ over the variables indexed by V and $\mathcal{G} = (V, E)$ the graph with respect to which $I_f(\cdot, \cdot | \cdot)$ has the Pairwise Markov property (P). Let $x^* \in \mathcal{X}$ be an arbitrary, but fixed, element of $\mathcal{X}$. We define for all $A \subseteq V$ the function

$$g_A(x) = f(x_A, x^*_{A^C})$$

where $(x_A, x^*_{A^C})$ is an element y with $y_\gamma = x_\gamma$ if $\gamma \in A$ and $y_\gamma = x^*_\gamma$ if $\gamma \notin A$. Since x^* is fixed, g_A depends on x through x_A only. Now, for all $A \subseteq V$, let

$$f_A(x) = \sum_{B : B \subseteq A} (-1)^{|A \backslash B|} g_B(x)$$

This formula implies that $f_A(x)$ depends on x through x_A only.

By applying the Moebius inversion lemma to the functions f and g we get:

$$f(x) = g_V(x) = \sum_{A : A \subseteq V} f_A(x)$$

We will show next that $f_A(x) \equiv 0$ whenever A is not a clique of $\mathcal{G}$. This fact, along with absorbing f_A into f_M whenever A is not a maximal clique and where $A \subset M \in MaxCliques(\mathcal{G})$, will prove our factorization property (F) over the maximal cliques of the graph $\mathcal{G}$. (absorption: if

$A \subset M \in MaxCliques(\mathcal{G})$ we can redefine $f'_M(x) = f_M(x) + f_A(x)$ and $f'_A(x) \equiv 0$).

To show that $f_A(x) \equiv 0$ whenever A is not a clique of $\mathcal{G}$, let $\alpha, \beta \in A$ such that $(\alpha, \beta) \notin E$ and let $C = A \backslash \{\alpha, \beta\}$. Then we have

$$f_A(x) = \sum_{B: B \subseteq C} (-1)^{|C \backslash B|} \quad \{ \quad g_B(x) - g_{B \cup \{\alpha\}}(x) - g_{B \cup \{\beta\}}(x) + g_{B \cup \{\alpha,\beta\}}(x)\}$$

We now want to show that $g_B(x) - g_{B\cup\{\alpha\}}(x) - g_{B\cup\{\beta\}}(x) + g_{B\cup\{\alpha,\beta\}}(x) \equiv 0$ for all $B \subseteq C = A\backslash\{\alpha, \beta\}$, which will prove our claim. $(\alpha, \beta) \notin E$ implies that $I_f(\alpha, \beta | V\backslash\{\alpha, \beta\})$ (by (P)), so there exist f_1, f_2 such that

$$f(V) = f_1(\alpha, V\backslash\{\alpha, \beta\}) + f_2(\beta, V\backslash\{\alpha, \beta\})$$

i.e.,

$$f(x_V) = f_1(x_\alpha, x_{V\backslash\{\alpha,\beta\}}) + f_2(x_\beta, x_{V\backslash\{\alpha,\beta\}}) \ \forall x_V \in \mathcal{X}$$

by considering x_V of the form $(x_B, x_\alpha, x_\beta, x^*_R) \ \forall x_B \in \mathcal{X}_B, x_\alpha \in \mathcal{X}_\alpha, x_\beta \in \mathcal{X}_\beta$ where $R = V\backslash(B \cup \{\alpha, \beta\})$ we get

$$\begin{aligned} g_{B\cup\{\alpha,\beta\}}(x) &= f(x_B, x_\alpha, x_\beta, x^*_R) \\ &= f_1(x_B, x_\alpha, x^*_R) + f_2(x_B, x_\beta, x^*_R) \quad (f1) \end{aligned}$$

for all $x_B \in \mathcal{X}_B, x_\alpha \in \mathcal{X}_\alpha, x_\beta \in \mathcal{X}_\beta$. By instantiating x_β in the formula $(f1)$ with x^*_β we get

$$\begin{aligned} g_{B\cup\{\alpha\}}(x) &= f(x_B, x_\alpha, x^*_\beta, x^*_R) \\ &= f_1(x_B, x_\alpha, x^*_R) + f_2(x_B, x^*_\beta, x^*_R) \end{aligned}$$

for all $x_B \in \mathcal{X}_B, x_\alpha \in \mathcal{X}_\alpha$. Similarly by instantiating x_α in in the formula $(f1)$ with x^*_α we get

$$\begin{aligned} g_{B\cup\{\beta\}}(x) &= f(x_B, x^*_\alpha, x_\beta, x^*_R) \\ &= f_1(x_B, x^*_\alpha, x^*_R) + f_2(x_B, x_\beta, x^*_R) \end{aligned}$$

for all $x_B \in \mathcal{X}_B, x_\beta \in \mathcal{X}_\beta$. And finally, by instantiating both x_α and x_β with x^*_α and x^*_β respectively, in the formula $(f1)$ we get

$$\begin{aligned} g_B(x) &= f(x_B, x^*_\alpha, x^*_\beta, x^*_R) \\ &= f_1(x_B, x^*_\alpha, x^*_R) + f_2(x_B, x^*_\beta, x^*_R) \end{aligned}$$

for all $x_B \in \mathcal{X}_B$. Now computing the formula $(*) = g_B(x) - g_{B\cup\{\alpha\}}(x) - g_{B\cup\{\beta\}}(x) + g_{B\cup\{\alpha,\beta\}}(x)$ with these alternative expansions we get

$$\begin{aligned} (*) &= f_1(x_B, x^*_\alpha, x^*_R) + f_2(x_B, x^*_\beta, x^*_R) \\ &\quad -f_1(x_B, x_\alpha, x^*_R) - f_2(x_B, x^*_\beta, x^*_R) \\ &\quad -f_1(x_B, x^*_\alpha, x^*_R) - f_2(x_B, x_\beta, x^*_R) \\ &\quad +f_1(x_B, x_\alpha, x^*_R) + f_2(x_B, x_\beta, x^*_R) \\ &= 0 \end{aligned}$$

■

In the particular case when f is a probability distribution the last implication in the previous theorem ((P) →(F)) is known as the Hammersly-Clifford theorem [Besag '74]. See also [Lauritzen '96, Jordan '05]. To the best of our knowledge, the proof described above is the first proof which holds for the general case of conditional independence with respect to a function f which takes values into an Abelian Group.

4 Particular Cases

4.1 Probability Theory

In the case when we consider functions $f : \mathcal{X} \to (0, \infty)$ where the group is $((0, \infty), \cdot, 1, \ ^{-1})$ and additionally we impose the constraint that $\sum_{x \in X} f(x) = 1$, or more generally $\int_X df = \int_X f(x)dx = 1$, we obtain strictly positive probability distributions and the notion of conditional independence becomes probabilistic conditional independence. By the factorisation theorem with respect to an associated graph $\mathcal{G}$ we can decompose the probability distribution in terms of clique potentials f_C as:

$$f(V) = \prod_{C \in MaxCliques(\mathcal{G})} f_C(C)$$

This is precisely the Hammersly-Clifford theorem [Besag '74, Lauritzen '96, Jordan '05] .

4.2 Additive Decomposability / Value Theory

In the case when we consider functions $f : \mathcal{X} \to \mathbb{R}$ where the group is $(\mathbb{R}, +, 0, -)$ we obtain an additive decomposition of the function f over the maximal cliques of the associated graph $\mathcal{G}$.

$$f(V) = \sum_{C \in MaxCliques(\mathcal{G})} f_C(C)$$

This decomposition theorem can be used to decompose value functions or fitness functions (as in Genetic Algorithms fitness). A set of theorems in the same spirit, while not in the same framework are the utility decomposition theorems. See [Bacchusassociated and Grove '95] and references thereof.

4.3 Relational Algebra

A relation is a function $r : \mathcal{X} \to \{0, 1\}$. If we consider the group $\mathbb{Z}_2 = (\{0, 1\}, \otimes, 0, -)$ we can decompose any relation r in terms of smaller relations defined over subsets of V. In this case the factorization theorem with respect to an associated graph $\mathcal{G}$ will be:

$$r(V) = \bigotimes_{C \in MaxCliques(\mathcal{G})} r_C(C)$$

5 Conclusion and Discussion

In this paper we have introduced a very general notion of Conditional Independence/Decomposability (section 2). Following the intuitions derived from the general case we introduced a notion of Conditional Independence relative to a function f which takes values into an Abelian Group $I_f(\cdot,\cdot|\cdot)$. We then proved that $I_f(\cdot,\cdot|\cdot)$ satisfies the following four properties: trivial independence, symmetry, weak union and intersection, which are cherished as essential/defining properties for notion of independence [Pearl '88]. As a consequence of these properties holding we obtained the equivalence of the Global, Local and Pairwise Markov Properties for our notion conditional independence relation $I_f(\cdot,\cdot|\cdot)$ [Pearl and Paz '87]. We proved the main theorem of this paper, which allows us to factorize the function f over the cliques of an associated Markov Network which reflects the Conditional Independences of subsets of variables with respect to f. This theorem is the natural generalisation of the Hammersly-Clifford theorem, for the more general case of functions that take values into an Abelian Group. The theory developed in this paper subsumes: probability distributions, additive decomposable functions and relations, as particular cases of functions over Abelian Groups:.

In comparison with the more traditional framework for probabilistic independence i.e., graphoids [Pearl '88] our notion of independence stands as follows: We do not support two axioms: namely, contraction $[I(A,D|C)\ \&\ I(A,B|C\cup D)\ \Rightarrow\ I_f(A,B\cup D|C)]$ and decomposition $[I(A,B\cup D|C)\Rightarrow I(A,B|C)]$. Decomposition is essentially a way to support non-saturated independence statements and hence it requires marginals in the model, but marginals are not generally available so it understandable that we have to drop it in our quest for generality. Contraction had to be dropped too because one of its premises contains a non-saturated statement. A weaker version however, called Weak Contraction $[I_f(A\cup B,D|C)\ \&\ I_f(A,B|C\cup D)\Rightarrow I_f(A,B\cup D|C)]$ that was used more recently in [Geiger and Pearl '93] does not contains non-saturated statements and is satisfied by our independence relation $I_f(\cdot,\cdot|\cdot)$. This can easily be shown because Weak Union and Intersection imply Weak Contraction.

[Geiger and Pearl '93] show that the saturated trivial independence, symmetry, weak union and intersection axioms are equivalent with graph separability. They also prove for this case a completeness theorem with respect to positive probability distributions. We were able to generalize this completeness theorem for the case of functions taking values into an Abelian group. (see Appendix / [Silvescu and Honavar '05]). This is one more piece of evidence in support of the fact that these four properties in the saturated setup, as our case also is, are indeed an axiomatic core for Independence.

References

Bacchus, F. and Grove, A. (1995). Graphical Models for Preference and Utility. In *Uncertainty in Artificial Intelligence*, 11.

Besag, J. (1974) Spatial Interaction and Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B (Methodological),* **36**(2):192-236.

Boutilier, C., Friedman, N., Goldszmidt M. and Koller, D. (1996) Context specific independence in Bayesian networks *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence (UAI-96)*, 115-123.

Cowell, R,. Dawid, P., Lauritzen, S. and Spiegelhalter, D. (1999) Probabilistic Networks and Experts Systems. Springer.

Geiger, D. and Pearl, J. (1993). Logical and Algorithmic Properties of Conditional Independence and Graphical Models. *The Annals of Statistics*, **21**(4):2001-2021.

Jensen, F., (2001) Bayesian Networks and Decision graphs. Springer.

Jordan, M. (2005). An Introduction to Probabilistical Graphical Models. *(to appear)*.

Lauritzen, S. (1989). Lectures on contingency tables. *Technical Report R 89-29*, Institute for Electronic Systems. Aalborg University.

Lauritzen, S. (1996). Graphical Models. Oxford Science Publications.

Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann Publishers.

Pearl, J. and Paz, A. (1987) Graphoids: a graph based logic for reasoning about relevancy relations. In *Advance in Artificial Intelligence-II,* 357-363. North-Holland.

Silvescu, A. and Honavar, V. (2005). On Independence and the Decomposability of functions taking values into Abelian Groups. (longer version of this paper).

www.cs.iastate.edu/~silvescu/papers/idfab/idfab.pdf

Studeny, M. (1992) Conditional independence relations have no finite complete characterization. In *Transactions if the 11th Prague conference on information theory, statistical decision functions and random processes,* 377-396.

6 Appendix

6.1 Markov properties equivalence proof

We will prove that (G) => (L) => (P) => (G).

(G) => (L) is trivial by taking $A = \{\alpha\}$ and $B = V\backslash(\{\alpha\}\cup \mathcal{N}(\alpha)))$.

(L) => (P) We want to prove that $\forall\alpha \in V \quad I(\alpha, V\backslash(\{\alpha\} \cup \mathcal{N}(\alpha)))|\mathcal{N}(\alpha))$ implies that $(\alpha,\beta) \notin E \Rightarrow I(\alpha,\beta|V\backslash\{\alpha,\beta\})$. Since $(\alpha,\beta) \notin E$ implies$\beta \notin \mathcal{N}(\alpha)$ it follows that $\beta \in V\backslash(\{\alpha\} \cup \mathcal{N}(\alpha)))$. By using the weak union property with $A = \{\alpha\}$, $B = \{\beta\}$, $C = \mathcal{N}(\alpha)$,$D = V\backslash(\{\alpha\} \cup \{\beta\} \cup \mathcal{N}(\alpha)))$ and the fact that $(\alpha,\beta) \notin E \Rightarrow \beta \notin \mathcal{N}(\alpha)$ it follows that $I(\alpha,\beta|\mathcal{N}(\alpha) \cup (V\backslash(\{\alpha\} \cup \{\beta\} \cup \mathcal{N}(\alpha))))$ or equivalently $I(\alpha,\beta|V\backslash(\{\alpha\} \cup \{\beta\}))$ q.e.d.

(P) => (G) we want to prove that: (Prop) if C separates A and B in the graph $\mathcal{G}$, then $I(A,B|C)$

We will prove the property (Prop) by induction after cardinality of the separator C, $|C|$.

Base case: For the case when $|C| = |V| - 2$ the property follows trivially from the from the Pairwise Markov Property (P). Also for $|C| > |V| - 2$ all the independence statements are trivial because at least one of A or B is empty; and since $I(\cdot,\cdot|\cdot)$ satisfies the trivial independence property they are true by default.

Inductive case: We will assume that the property is true for $|C| \geq k$ and prove it for $|C| = k - 1$, where $1 \leq k \leq |V| - 2$. Let $|C| = k - 1$, then we want to prove that if C separates A and B in the graph $\mathcal{G}$, then $I(A,B|C)$. Since $k \leq |V| - 2$ it follows that $|C| = k - 1 \leq |V| - 3$ Hence at least one of the sets A and B has at least two elements. Because of the symmetry property we can assume without loss of generality that the set is A. Let $\alpha \in A$. From the fact that C separates A and B in the graph $\mathcal{G}$ it follows that $C \cup \{\alpha\}$separates $A\backslash\{\alpha\}$ from B in $\mathcal{G}$. Using the inductive assumption we get that $I(A\backslash\{\alpha\},B|C \cup \{\alpha\})$. Also from the fact that C separates A and B in the graph $\mathcal{G}$ it follows that $C \cup (A\backslash\{\alpha\})$separates $\{\alpha\}$ from B in $\mathcal{G}$ and again using the inductive assumption we get that $I(\alpha,B|C \cup (A\backslash\{\alpha\}))$. Now by the intersection property we have that $I(A\backslash\{\alpha\},B|C \cup \{\alpha\})$ & $I(\alpha,B|C \cup (A\backslash\{\alpha\})) \Rightarrow I(A,B|C)$.

6.2 Proof of the Moebius Inversion Lemma

(2) => (1)

$$\begin{aligned}\sum_{B:B\subseteq A} f(B) &= \sum_{B:B\subseteq A}\sum_{C:C\subseteq B}(-1)^{|B\backslash C|}g(C)\\ &= \sum_{C:C\subseteq A}\sum_{B:C\subseteq B\subseteq A}(-1)^{|B\backslash C|}g(C)\\ &= \sum_{C:C\subseteq A}\sum_{D:D\subseteq A\backslash C}(-1)^{|D|}g(C)\end{aligned}$$

$\sum_{D:D\subseteq A\backslash C}(-1)^{|D|}g(C)$ is 0 unless $A\backslash C = \emptyset$, or equivalently $A = C$ (because $C \subseteq A$), and therefore the whole sum will be $g(A)$, which proves the desired equality. To see that if $E = A\backslash C \neq \emptyset$ then $\sum_{D:D\subseteq E}(-1)^{|D|}g(C)$ is 0 we observe that there are as many pluses in this sum as there are minuses and therefore the terms cancel out. $(0 = (1-1)^{|D|} = \sum_{D:D\subseteq E}(-1)^{|D|}$ if $D \neq \emptyset$; if $D = \emptyset$ the sum is by definition $0^0 = 1)$

(1) => (2)

$$\begin{aligned}\sum_{B:B\subseteq A}(-1)^{|A\backslash B|}g(B) &= \sum_{B:B\subseteq A}\sum_{C:C\subseteq B}(-1)^{|A\backslash B|}f(C)\\ &= \sum_{C:C\subseteq A}\sum_{B:C\subseteq B\subseteq A}(-1)^{|A\backslash B|}f(C)\\ &= \sum_{C:C\subseteq A}\sum_{D:D\subseteq A\backslash C}(-1)^{|D|}f(C)\end{aligned}$$

Which by the same argument as before is equal to $f(A)$, which proves the desired equality.

6.3 Graph Separability and Independence

Definition (associated dependence graph): Given a set Σ of pairwise conditional independence statements $I(\alpha,\beta|V\backslash\{\alpha,\beta\})$, a graph $\mathcal{G}(\Sigma) = (V,E)$ is called the associated dependence graph if $(\alpha,\beta) \notin E \Leftrightarrow I(\alpha,\beta|V\backslash\{\alpha,\beta\}) \in \Sigma$.

Definition (closure): Let $(X_\alpha)_{\alpha\in V}$ be a collection of variables indexed by V , Σ be an arbitrary set of independence statements of the form $I(A,B|C)$, where A,B,C is a partition of V, and $\mathcal{A}$ a set of axioms. We denote by Σ^+ the set of all independence statements that can be inferred from the independence statements in Σ in a finite number of steps by using only axioms from the set $\mathcal{A}$. If such is the case, we call Σ^+ *the closure of* Σ *under the axioms* $\mathcal{A}$.■

In this section we will show that Conditional Independence with respect to a function f and graph separation in the associated graph are identical concepts.

Theorem (separability $\Leftrightarrow$ conditional independence): [Geiger & Pearl '93] *Let Σ be a set of saturated independence statements over a finite set of variables $(X_\alpha)_{\alpha\in V}$ indexed by elements from V. Let Σ^+ be the closure of Σ with respect to saturated trivial independence, symmetry, intersection and weak union. And let $\mathcal{G}(\Sigma^+)$ the dependence graph associated with set of pairwise independence statements in Σ^+. Then for any A,B,C partition of V we have:*

$$Sep_{\mathcal{G}}(A,B|C) \Leftrightarrow I(A,B|C) \in \Sigma^+$$

Proof. See [Geiger & Pearl '93] Theorem 13 for a proof of this theorem. ■

Corollary: $I_f(\cdot,\cdot|\cdot)$ *satisfies the equivalence between graph separability and Independence stated in the previous theorem.*

6.4 Completeness

Definition (satisfaction): Let $f : \mathcal{X} \to G$ be a function that takes values into an Abelian Group $(G, +, 0, -)$ and let A, B, C is a partition of V. We say that f satisfies a conditional independence statement $\sigma = I(A, B|C)$ if $I_f(A, B|C)$.

Definition (Markov Network): A graph $\mathcal{G} = (V, E)$ is called a Markov Network for a function $f : \mathcal{X} \to G$ iff for all partitions into three sets of V, A, B, C we have $Sep_{\mathcal{G}}(A, B|C) \Rightarrow I_f(A, B|C)$.

Theorem (completeness): *Let $f : \mathcal{X} \to G$ be a function that takes values into an Abelian Group $(G, +, 0, -)$. Let Σ be a set of saturated independence statements over sets in V, and let Σ^+ be the closure of Σ with respect to trivial independence, symmetry, intersection and weak union. Then for every $\sigma \notin \Sigma^+$, there exists a function $f : \mathcal{X} \to G$ such that f satisfies Σ^+ and does not satisfy σ.*

Proof. By the previous theorem there exists a graph $\mathcal{G}(\Sigma^+)$ (the associated graph) that satisfies Σ^+ and no other independence statement. So to prove the theorem we need to show that there exists a function $f : \mathcal{X} \to G$ that satisfies the statements that hold in $\mathcal{G}(\Sigma^+)$ and does not satisfy σ. We will prove this in the next lemma. ■

Lemma. *Let $\mathcal{G} = (V, E)$ a graph and let A, B, C a partition of V such that C does not separate A from B in the graph $\mathcal{G}$. Let $(G, +, 0, -)$ a group with at least two elements. Then there exists a function $f : \mathcal{X} \to G$ such that $\mathcal{G}$ is a Markov Network for f and $\neg I_f(A, B|C)$.*

Proof. Let $(X_\alpha)_{\alpha \in V}$ be a collection of variables indexed by V such that all variables X_α have the same range $\{a, b\}$. We will construct a function $f : \mathcal{X} = \times_\alpha \mathcal{X}_\alpha \to G$ such that $\mathcal{G}$ is a Markov Network for f and $\neg I_f(A, B|C)$. Because G is a group with at least two elements, let $g \in G$ such that $g \neq 0$. From the fact that C does not separate A from B in the graph $\mathcal{G}$ and the fact that A, B, C is a partition of V it follows that there exists a $\alpha \in A$ and $\beta \in B$ such that $(\alpha, \beta) \in E$. Let

$$f(x_V) = h(x_\alpha, x_\beta) \sum_{x_\gamma \in V \setminus \{\alpha, \beta\}} g(x_\gamma)$$

where $g(x) = 0 \; \forall x \in \{a, b\}$and

$$h(x, y) = \begin{cases} 0 & if \quad (x, y) = (a, a) \\ 0 & if \quad (x, y) = (a, b) \\ 0 & if \quad (x, y) = (b, a) \\ g & if \quad (x, y) = (b, b) \end{cases}$$

Note that $h(x, y)$ is not decomposable into smaller functions (i.e., $h(x, y) \neq h_1(x) + h_2(y)$- see a proof next). The function f has the desired properties, that is $I_f(\alpha, \beta|V \setminus \{\alpha, \beta\})$ does not hold because $h(x, y)$ is not decomposable and $f(x_V) = h(x_\alpha, x_\beta)$ hence f cannot be decomposed into functions that separate x_α and x_β. Furthermore, $Sep_{\mathcal{G}}(A, B|C) \Rightarrow I_f(A, B|C)$ because $I_f(A, B|C)$ holds for all sets A, B that do not separate α and β (easy to check).

We will now prove that $h(x, y)$is not decomposable by contradiction. Assume $h(x, y) = h_1(x) + h_2(y)$. Then

$$h(a, a) = 0 = h_1(a) + h_2(a) \quad (1)$$

$$h(a, b) = 0 = h_1(a) + h_2(b) \quad (2)$$

$$h(b, a) = 0 = h_1(b) + h_2(a) \quad (3)$$

$$h(b, b) = g = h_1(b) + h_2(b) \quad (4)$$

from equations (1) and (2) it follows (by simplifying $h_1(a)$) that $h_2(a) = h_2(b)$; Similarly from equations (1) and (3) it follows (by simplifying $h_2(a)$) that $h_1(a) = h_1(b)$. Let $g_1, g_2 \in G$ such that $g_1 = h_2(a) = h_2(b)$ and $g_2 = h_1(a) = h_1(b)$. Then from equation (1) it follows that $g_1 + g_2 = 0$, and from equation (4) it follows that $g_1 + g_2 = g$, hence $g = 0$, contradiction.

This completes the proof of the lemma.■

Note that if the group G has only one element then there exists only one function from any domain to G (the one that maps all the elements of the domain to the unique element in G) and therefore the problem of completeness is vacuous.

The **(completeness)** theorem along with the **(separability $\Leftrightarrow$ conditional independence)** theorem allow us to determine conditional independence statements relative to the function f by the means of examining separation in the associated graph.

6.5 Intersection + Weak Union => Weak Contraction

Proposition (INT + WU => WC): *INT + WU => WC*

Proof.

WC: $I(A \cup B, D|C)$ & $I(A, B|C \cup D) \Rightarrow I(A, B \cup D|C)$

INT: $I(A, D|C \cup B)$ & $I(A, B|C \cup D) \Rightarrow I(A, B \cup D|C)$

Assume $I(A \cup B, D|C)$ (1) and $I(A, B|C \cup D)$ (2)

$I(A \cup B, D|C) \Rightarrow_{WU} I(A, D|C \cup B)$ (3)

(2) & (3) $\Rightarrow_{INT} I(A, B \cup D|C)$■

6.6 C,D and WC

1. **C:** $I(A, D|C)$ & $I(A, B|C \cup D) \Rightarrow I_f(A, B \cup D|C)$
2. **D:** $I(A, B \cup D|C) \Rightarrow I(A, B|C)$
3. **WC:** $I_f(A \cup B, D|C)$ & $I_f(A, B|C \cup D) \Rightarrow I_f(A, B \cup D|C)$