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 relevant variables of the problem under consideration. In this paper we investigate the decomposability of functions which take values into an Abelian Group. Examples of such functions include: probability distributions, energy functions, value functions, fitness functions, and relations. For such problems we define a notion of conditional independence between subsets of the problem's variables. We prove a decomposition theorem that relates independences between subsets of the problem's variables with a factorization property of the respective function. As particular cases of this theorem we retrieve the Hammersley-Clifford theorem for probability distributions; an Additive Decomposition theorem for energy functions, value functions, fitness functions; and a relational algebra decomposition theorem.

1 Introduction

Probabilistic Graphical Models (PGMs) have proved to be an effective 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 idea behind PGMs is the notion of probabilistic independence. Independence allows us to "decompose" the probability distribution into smaller parts, thereby substantially reducing the number of independent parameters that we need to know in order to specify the probability distribution.

Given the successful exploitation of independence in PGMs it is natural to ask whether we can define a more general case of independence with similar decomposability properties. Results concerning additive (rather than multiplicative) decomposability in various scenarios have been explored in [Bertelle and Brioschi '72] [Keeney and Raiffa '76] [Bacchus and Grove '95]. Similarly, in the context of relational databases decomposition results such as MVD (Multi-Valued Decomposition) were obtained, see [Butz '00] and references therein.

In this paper, we seek a unifying thread of decomposability results. More precisely we 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 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 General 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 precisely, we prove a generalization of the well known Hammersley-Clifford theorem, which holds for arbitrary functions from $\mathcal{F}_{\cdot\to AG}$.

These results, in addition to unifying a broad class of decomposition results that have been previously proved in particular cases, should hopefully clear the path for porting results and techniques developed in settings that represent particular cases of $\mathcal{F}_{\cdot\to AG}$.

The rest of the paper is organized as follows: In section 2 we introduce some definitions regarding Decomposition and Independence in a general setting. In section 3 we shrink our domain of interest and define a General Conditional Independence concept for functions whose ranges are Abelian Groups: $\mathcal{F}_{\cdot\to AG}$. Subsequently, we explore some of the properties of this generalized independence relation. This section also contains the main result of this paper: a factorization property consisting in the natural generalization of the Hammersley-Clifford theorem [Besag '74] for arbitrary functions from $\mathcal{F}_{\cdot\to AG}$. In section 4 we present some important particular cases of decomposable functions which take values into an Abelian Group, such as: probability distributions, additive decomposable functions and relations. Section 5 concludes with a summary, discussion, and a brief outline of some directions for further research.

2 Decomposition and Independence

Decomposition is an key technique that makes the solution of otherwise complex problems tractable by the means of a divide and conquer approach. . Decomposition exploits the fact that occasionally a problem can be split (*decomposed*) into subproblems 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 subproblems, 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 more of the subproblems are further decomposable into smaller parts, we can apply the same 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.

In what follows we will try to capture these intuitions underlying Decomposition and Independence 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 - 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 isomorphic 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 $\oplus_P^{P_1,P_2} : S_{AC} \times S_{BC} \to S_D$. ■

The rest of this paper is concerned with variable-based independence. More precisely, we will consider the particular case where 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 yielding a single operator: $\oplus : G \times G \to G$. Lastly, we will consider G to be an Abelian Group for reasons that will become apparent later on.

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

In this section we consider Independence and Decomposability of functions whose ranges are Abelian Groups: $\mathcal{F}_{\cdot \to AG}$. After defining Abelian Groups (3.1), we introduce the definition of Conditional Independence with respect to a function f, - $I_f(\cdot, \cdot|\cdot)$ (3.2) (Note that for probabilistic independence the function f is the very probability distribution). We then show that $I_f(\cdot, \cdot|\cdot)$ satisfies four properties that are considered as essential/defining for the notion of independence, see [Pearl '88] (3.3). These properties are: trivial independence, symmetry, weak union and intersection. We then recapitulate the main results already existing in the literature, (e.g., [Geiger and Pearl '93]), regarding Conditional Independence relations that satisfy these four properties (3.4). The main target of these results is to establish the equivalence between conditional independence and graph separability (Note: the graph in question is obtained by not drawing an edge between two variables whenever they are independent of each other given the rest of the variables, and drawing one otherwise). We then present the main theorem which allows us to "piece down" a set of pairwise conditional indepeindependencesdences of the form $I_f(A, B|C)$ into a global decomposition of the function f over the maximal cliques of the associated graph. This theorem is the natural generalization of the Hammersley-Clifford theorem for probability distributions, to the more general case of $\mathcal{F}_{\cdot \to AG}$

3.1 Abelian Groups

We start with the definition of Abelian Groups (a.k.a. commutative 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 $\forall g \in G\ \exists! \ominus g \in G\ s.t\ g \oplus \ominus g = \ominus g \oplus g = \theta$ ■

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 shorthand 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

We now proceed to define a general notion of Conditional Independence with respect to a function f, $I_f(\cdot, \cdot|\cdot)$. This independence relation is basically a formalization of the intuition behind decomposition presented in Section 2: namely, that independence should allow us to decompose a problem into subproblems, solve them separately and then combine the results. We start with some notations, then present the definition and 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 sets of variable indices $A, B, C \subseteq V$ we will assert conditional independence statements regarding the associated subsets of variables X_A,X_B,X_C such as: the sets of variables X_A and X_B are independent conditioned on X_C and write $I(X_A, X_B|X_C)$. We will actually use the shorthand formulation, A and B are independent conditioned on C, 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 an Abelian 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 we 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 $A \subsetneqq V$ 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.

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, see [Lauritzen '96]. 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, Cowell et. al. ' 99] have identified as desirable for any conditional independence relation because 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, thus providing supporting evidence that this is the "right" concept .

Theorem 1 (independence properties): *Let $(G,+,0,-)$ be an Abelian 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. See the extended version of the paper [Silvescu and Honavar '05].■

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 start with a survey some known results regarding Conditional Independence relations satisfying the above-mentioned four properties [Geiger and Pearl '93]. We first introduce some graph terminology, then define different types of Markov properties and subsequently, establish their equivalence. We end with a theorem that states the equivalence between graph separability and conditional independence. All results hold under the assumptions of: trivial independence, symmetry, weak union and intersection.

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, $\mathcal{N}(A)=\{\beta|\beta\notin A\ and\ \exists\alpha\in A\ such\ that\ (\alpha,\beta)\in E\}$.

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 $\mathcal{G} = (V, E)$.

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 2 (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 still holds. See the longer version of this paper [Silvescu & Honavar '05] for a complete proof. ■

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

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* $\mathcal{A}$.■

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 $(\alpha, \beta) \notin E \Leftrightarrow I(\alpha, \beta|V \backslash \{\alpha, \beta\}) \in \Sigma$. In genera,l given a set Σ of not necessarily pairwise conditional independence statements we define the set $\Sigma_{pairwise}$ as the set of of all pairwise independence statements that can be inferred from Σ using the ***Trivial independence, Symmetry, Weak union*** *and* ***Intersection*** axioms (i.e., all pairwise Independence statements from the closure of Σ, - Σ^+). Furthermore we define the associated dependence graph $\mathcal{G}(\Sigma)$ of such a general set of conditional independence statements Σ as the associated dependence graph of $\Sigma_{pairwise}$. ■

Theorem 3 (separability $\Leftrightarrow$ conditional independence): [Geiger and 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:*

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

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

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

So far we have seen that any set Σ of Conditional Independence statements produces a graph $\mathcal{G}(\Sigma^+)$ such that separability in this graph is equivalent to Conditional Independence in the closure of Σ. If the Independence relation in question is $I_f(\cdot, \cdot|\cdot)$ we have furthermore that $Sep_{\mathcal{G}(\Sigma^+)}(A, B|C) \Leftrightarrow I_f(A, B|C) \in \Sigma^+ \Leftrightarrow f(A, B, C) = f_1(A, C) + f_2(B, C)$. We will next prove the main result of the paper, namely, a theorem that will allow us to "compile" pairwise decompositions that are implied by conditional independence statements between two sets of variables conditioned on a third one, of the form $I_f(A, B|C) \in \Sigma^+ \Rightarrow f(A, B, C) = f_1(A, C) + f_2(B, C)$ into a "finer" decomposition over the maximal cliques of the associated graph $\mathcal{G}(\Sigma^+)$. In other words, this theorem shows that if the four properties are satisfied, we can "boil down" a set of pairwise decompositions to one "wholisticholistic" decomposition over the maximal cliques of the associated graph $\mathcal{G}(\Sigma^+)$.

3.5 The factorization theorem

We now proceed to prove the 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 associated 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 $f : \mathcal{X} \rightarrow G$ 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 there exist a collection of functions $\{f_C : \mathcal{X}_C \rightarrow G\}_{C \in MaxCliques(\mathcal{G})}$

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

Theorem 4 (factorization): *Let $\mathcal{G} = (V, E)$ be an undirected graph, $(G, +, 0, -)$ be an Abelian Group, $(X_\alpha)_{\alpha \in V}$ be a collection of variables indexed by V, $f : \mathcal{X} \rightarrow G$ be a function from the set $\mathcal{X}$ to the group G. Let $I_f(\cdot, \cdot|\cdot)$ conditional independence relation induced by the function f . Then* (G) $\Leftrightarrow$ (L) $\Leftrightarrow$ (P) $\Leftrightarrow$ (F), *where all the Markov properties are with respect to $I_f(\cdot, \cdot|\cdot)$.* ■

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 from the ***Markov properties*** theorem in the previous section.

(F) $\Rightarrow$ (G) Let $\mathcal{G} = (V, E)$ be a graph and $f : \mathcal{X} \rightarrow G$ 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 [Griffeath'76, Lauritzen '96, Jordan '05]. See also the longer version of this paper [Silvescu & Honavar '05]. ■

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

Let $f : \mathcal{X} \rightarrow G$ 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

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

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 $B \subseteq C = A\setminus\{\alpha,\beta\}$, which will prove our claim. $(\alpha,\beta) \notin E$ implies that $I_f(\alpha,\beta|V\setminus\{\alpha,\beta\})$ (by (P)), so there exist f_1, f_2 such that

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

i.e.,

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

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 $R = V\setminus(B \cup \{\alpha,\beta\})$ 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 Hammersley-Clifford theorem [Besag '74]. The proof technique based on the Moebius Inversion Lemma was first used for proving the Hammersley-Clifford theorem for probabilities in [Griffeath '76] , see also [Lauritzen '96, Jordan '05]. To the best of our knowledge, the proof presented is the first proof that holds for the general case of conditional independence with respect to a function f which takes values into an Abelian Group.

4 Particular Cases

We now review some important examples of functions over particular Abelian Groups and the associated factorization theorems.

Probability Theory In the case when we consider functions $f : \mathcal{X} \rightarrow (0,\infty)$ 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 factorization 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 Hammersley-Clifford theorem [Besag '74, Griffeath '76, Lauritzen '96, Jordan '05] .

Additive Decomposability / Value Theory When we consider functions $f : \mathcal{X} \rightarrow \mathbb{R}$ 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. A set of theorems in the same spirit, while not in the same framework are the utility decomposition theorems. See [Bacchus and Grove '95] and references therein.

Relational Algebra A relation is a function $r : \mathcal{X} \rightarrow \{0,1\}$. 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 Summary and Discussion

In this paper, we have introduced a general notion of Conditional Independence/Decomposability. Following the intuitions derived from the general case, we introduced the 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 held to be essential properties for the notion of independence [Pearl '88]. As a consequence, we obtained the equivalence of the Global, Local and Pairwise Markov Properties for $I_f(\cdot,\cdot|\cdot)$, as well as the equivalence between Conditional Independence and Graph Separability in the associated graph, based on well known results e.g., [Geiger and Pearl '93]. We then proved the main theorem of this paper, which allows us to "lift up" a set of pairwise Conditional Independences and consequently pairwise factorizations of the function f to a global factorization of f over the maximal cliques of the associated dependence graph. This theorem is the natural generalization of the Hammersley-Clifford theorem which holds for probability distributions, to the more general case of functions that take values into an Abelian Group. The theory developed in this paper subsumes: factorization of probability distributions, additive decomposable functions and decomposable relations, as particular cases of functions over Abelian Groups.

In contrast with the more traditional framework for probabilistic independence i.e., graphoids [Pearl '88], our notion of independence does not support 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, but marginals are not generally available since they appear to be an idiosyncratic feature of probability distributions. Thus it is understandable that we have to drop decomposition in our quest for generality. Contraction had to be dropped as well 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 in [Geiger and Pearl '93] does not require non-saturated statements and is satisfied by our independence relation $I_f(\cdot,\cdot|\cdot)$. This follows from the observation that Weak Union and Intersection imply Weak Contraction. The Intersection property is sometimes dropped from probabilistic independence axioms in order to obtain semi-graphoids [Pearl '88]. Both the graphoid and semi-graphoid sets of axioms however have been shown to be incomplete for non-saturated probabilistic conditional independence statements [Studeny '92]. In fact [Studeny '92] shows that any finite set of axioms is an incomplete characterization of non-saturated probabilistic conditional independence statements. In our case, since we use only saturated independence statements due to the unavailability of marginals in the general case, a completeness theorem can be proved. In a longer version of this paper [Silvescu and Honavar '05] we prove a completeness theorem that states that the axioms of *trivial independence, symmetry, intersection and weak union* are a complete characterisation of Conditional Independences for functions over an Abelian Group G . This theorem is a natural generalization, for functions over Abelian groups, of a completeness theorem for positive probability distributions and the four above-mentioned axioms in a saturated probabilistic conditional independence.(which was previously obtained by [Geiger and Pearl '93]). This is one more important piece of evidence in support of the fact that these four properties in the saturated setup, as our case is, are indeed an axiomatic core for Conditional Independence.

References

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

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

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

Griffeath, D. (1976) Introduction to Random Fields. In Kemeny, J., J. Snell and A. Knapp (Eds), *Denumerable Markov Chains*. New York: Springer-Verlag.

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

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/idfablong.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.