Structural imsets: an algebraic method for describing conditional independence structures*

Milan Studený
Institute of Information Theory and Automation
Pod vodárenskou věží 4, 18208 Prague, Czech Republic
studeny@utia.cas.cz

Abstract

The contribution recalls basic ideas of the method of structural imsets which is a non-graphical method for describing probabilistic conditional independence structures based on the use of special integral (= integer-valued) vectors. The attention is devoted to the concept of independence implication between structural imsets defined as the inclusion of their induced conditional independence structures. An algebraic characterization of independence implication is recalled and implementation aspects are analyzed in more details from a theoretical point of view.

Keywords: Conditional independence structure, structural imset, independence implication.

1 Introduction

The general topic of this paper is the description of *conditional independence* (= CI) *structures*. Describing these structures by various graphs whose nodes correspond to variables appear not to be satisfactory: graphs are not able to describe all CI structures induced by discrete probability measures (for justification see § 3.6 of [8]). Moreover, the groundless limitation to a restricted class of graphical models may lead to serious inferential errors in statistical learning procedures (for respective arguments see § 1.1 of [8]).

These observations motivated a non-graphical method for describing probabilistic CI structures, namely the method of *structural imsets.* The basic ideas of this approach were already sketched out in the middle of the 1990's [5]. However, further results have since been achieved which amended, supplemented and gave more precision to the original idea. The aim of an upcoming research monograph [8] is to present didactically the method of structural imsets in its present state; readers interested in an outline of the method can consult the paper [7].

Structural imsets are certain integral (= integer-valued) vectors whose components correspond to subsets of the respective set of variables N. The point is that structural imsets are able to describe every probabilistic CI structure induced by a discrete probability measure over N. There are other pay-offs of this approach: one of them is the ability to describe inclusion of CI structures, named *independence implication*, in algebraic terms, more specifically, through the arithmetic of integers.

Because the dimension of structural imsets grows exponentially with the number of variables $|N|$ one has to consider implementation tasks thoroughly. This paper deals with some theoretical questions related to computer implementation of this method. More specifically, it is devoted to the question of testing structural imsets and the question of independence implication verification.

* This work was supported by the grant GAČR n. 201/04/0393 and the grant GAAVČR n. A1075104.

The structure of the paper is as follows. In Section 2 basic concepts are recalled: the concept of a CI structure, the concept of a structural imset and a closely related concept of a *combinatorial imset*. In Section 3 a basic (direct) algebraic characterization of independence implication is recalled and an alternative (dual) one is mentioned. The core of the paper is Section 4 which deals with implementation aspects. It is explained in Sections 4.1 and 4.2 that the task of testing whether an imset u is structural can be transformed to the task of decomposing a certain multiple $n \cdot u$, $n \in \mathbb{N}$ of the imset into elementary factors and the maximal constant n only depends on the number of variables $|N|$. An analogous result for testing independence implication of structural imsets is mentioned in Section 4.3. Moreover, the question of how to get the maximal value of the respective multiplicative constant is discussed there. Section 5 is a response to a proposal raised by a reviewer. The paper is concluded by Section 6 where two other potential applications of the method of structural imsets are mentioned.

2 Basic concepts

Let ξ_i, $i \in N$ be a collection of discrete random variables indexed by a finite non-empty set N. Its distribution is a discrete measure P over N, that is, a probability measure on $\mathsf{X}_N \equiv \prod_{i\in N} \mathsf{X}_i$ where X_i, $i \in N$ are finite non-empty sets such that ξ_i takes values in X_i.

CI statements concerning ξ_i, $i \in N$ will correspond to triplet $\langle A, B|C\rangle$ of pairwise disjoint subsets of N: $A, B, C \subseteq N$, $A \cap B = A \cap C = B \cap C = \emptyset$. The class of these triplets will be denoted by $\mathcal{T}(N)$. If $A, B \subseteq N$ are disjoint then juxtaposition AB will be used to denote their union $A \cup B$. If $a \in N$ then the symbol a will also be used to denote the singleton $\{a\}$. Given $\langle A, B|C\rangle \in \mathcal{T}(N)$ the symbol $A \perp\!\!\!\perp B \,|\, C\ [P]$ will mean that the respective CI statement is valid, that is, $[\xi_i]_{i\in A}$ is conditionally independent of $[\xi_i]_{i\in B}$ given $[\xi_i]_{i\in C}$. A formal definition in terms of marginal densities of P is recalled here: for every $x \in \mathsf{X}_N$,

$$p_{ABC}(x_{ABC})\cdot p_C(x_C) = p_{AC}(x_{AC})\cdot p_{BC}(x_{BC}),$$

where p_A denotes the marginal density of P for $A \subseteq N$, which is a function on $\mathsf{X}_A \equiv \prod_{i\in A} \mathsf{X}_i$, and x_A is the projection of x onto X_A.

Definition 1 The *conditional independence structure* (CI structure) induced by P is the collection of triplets $\langle A, B|C\rangle \in \mathcal{T}(N)$ such that $A \perp\!\!\!\perp B \,|\, C\ [P]$. Any collection of triplets $\langle A, B|C\rangle \in \mathcal{T}(N)$ obtained in this way will briefly be called a (probabilistic) *CI structure over* N. ♢

CI structures are typically described by graphs over N, that is, graphs whose set of nodes is N. There are several types of graphs used for this purpose: undirected graphs, acyclic directed graphs and chain graphs (see [1]). This is done with the aid of a graphical criterion which allows one to decide whether a triplet $\langle A, B|C\rangle \in \mathcal{T}(N)$ is represented in a graph G over N. If a CI structure over N coincides with the set of triplets represented in G then G is used to describe the CI structure.

The drawback of graphical methods is that graphs are not able to describe all probabilistic CI structures because the number of different CI structures over N grows much faster with N than the number of graphs over N. This fact combined with the careless use of improper learning procedures may result in inferential errors as reported in § 1.2 of [8]. The limitation of graphical approaches was the main motive to introduce an algebraic method for describing probabilistic CI structures which uses certain integral vectors. Let $\mathcal{P}(N)$ denote the power set of N, that is, the class of all subsets of N.

Definition 2 An integer-valued function on $\mathcal{P}(N)$ will be called an *imset* over N. Given $\langle A, B|C\rangle \in \mathcal{T}(N)$ the respective imset $u_{\langle A,B|C\rangle}$ takes four non-zero values:

$$\begin{array}{rclcl} u_{\langle A,B|C\rangle}(ABC) &=& u_{\langle A,B|C\rangle}(C) &=& +1\,, \\ u_{\langle A,B|C\rangle}(AC) &=& u_{\langle A,B|C\rangle}(BC) &=& -1\,. \end{array}$$

An *elementary* imset is any imset $u_{\langle a,b|C\rangle}$ where $a, b \in N$, $a \neq b$ and $C \subseteq N \setminus ab$. The collection of elementary imsets over N will be denoted by $\mathcal{E}(N)$.

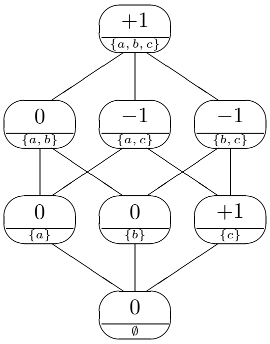

Figure 1: An elementary imset over $N = \{a, b, c\}$.

An imset u over N will be called *combinatorial* if it is a combination of elementary imsets with non-negative integral coefficients:

$$u = \sum_{v \in \mathcal{E}(N)} k_v \cdot v \quad \text{where } k_v \in \mathbb{Z}^+ .$$

The class of combinatorial imsets over N will be denoted by $\mathcal{C}(N)$.

An imset u over N will be called *structural* if it is a combination of elementary imsets with non-negative rational coefficients:

$$u : \mathcal{P}(N) \to \mathbb{Z}, \quad u = \sum_{v \in \mathcal{E}(N)} k_v \cdot v\,, \; k_v \in \mathbb{Q}^+ .$$

The class of structural imsets over N will be denoted by $\mathcal{S}(N)$. $\diamondsuit$

Note that the word **imset** is an abbreviation for **i**nteger-valued **m**ulti**set**. Imsets over a small set N can be visualized by special diagrams, namely the Hasse diagrams of $\mathcal{P}(N)$ labelled by numberical values – see Figure 1 for a diagram showing an elementary imset $u_{\langle a,b|c\rangle}$ over $N = \{a, b, c\}$.

Evidently, every elementary imset is combinatorial and every combinatorial imset is structural. Note that, at the present time, no example of a structural imset which is not combinatorial is known. Thus, it may be the case that $\mathcal{C}(N) = \mathcal{S}(N)$; it is true for $|N| \le 4$.

Structural imsets can describe probabilistic CI structures through an algebraic criterion. It is an analogue of a graphical criterion used in graphical modelling.

Definition 3 Let u be a structural imset over N and $\langle A, B|C\rangle \in \mathcal{T}(N)$. We say that $\langle A, B|C\rangle$ is *represented in* the imset u if there exists a natural number $k \in \mathbb{N}$ such that $k \cdot u - u_{\langle A,B|C\rangle} \in \mathcal{S}(N)$ and denote this by $A \perp\!\!\!\perp B \,|\, C\ [u]$. The set of triplets $\langle A, B|C\rangle \in \mathcal{T}(N)$ represented in $u \in \mathcal{S}(N)$ will be denoted by $\mathcal{M}_u$.

A discrete probability measure P over N will be called *Markovian* with respect to u if, for every $\langle A, B|C\rangle \in \mathcal{T}(N)$,

$$A \perp\!\!\!\perp B \,|\, C\ [u] \;\Rightarrow\; A \perp\!\!\!\perp B \,|\, C\ [P]\,.$$

If the converse implication also holds

$$A \perp\!\!\!\perp B \,|\, C\ [u] \;\Leftrightarrow\; A \perp\!\!\!\perp B \,|\, C\ [P]\,,$$

then P will be called *perfectly Markovian.* $\diamondsuit$

Thus, if a probability measure P is perfectly Markovian with respect to $u \in \mathcal{S}(N)$ then u can be used to describe the CI stucture induced by P. The point is that every probabilistic CI structure over N can be described in this way.

Theorem 1 Let P be a discrete probability measure over N. Then there exists a structural imset u over N such that P is perfectly Markovian with respect to u.

The above result follows from Theorem 5.2 in [8] which is more general: the same conclusion can be made even for some other classes of probability measures.

3 Independence implication

Another important concept is the concept of independence implication.

Definition 4 Let $u, v \in \mathcal{S}(N)$. We say that u *i-implies* v (here i- stands for independence) and write $u \rightharpoonup v$ if $\mathcal{M}_v \subseteq \mathcal{M}_u$. We say that u and v are *independence equivalent* if and write $u \rightleftharpoons v$ if $\mathcal{M}_u = \mathcal{M}_v$. $\diamondsuit$

Note that in the original series of papers [5] this implication was called ‘facial implication’ and denoted in another way. Clearly, one has

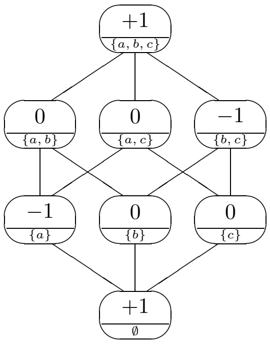

Figure 2: A combinatorial imset $u = u_{\langle a,bc|\emptyset\rangle}$.

$u \rightleftharpoons v$ iff $[u \rightharpoonup v$ and $v \rightharpoonup u]$. Thus, independence equivalence can be viewed as a special case of independence implication. In particular, the following direct algebraic characterization of independence implication can be applied to independence equivalence as well.

LEMMA 1 Given $u, v \in \mathcal{S}(N)$ one has $u \rightharpoonup v$ iff

$$\exists\, l \in \mathbb{N} \quad l \cdot u - v \in \mathcal{S}(N)\,. \tag{1}$$

If, moreover, $v \in \mathcal{C}(N)$, then it is equivalent to the condition

$$\exists\, k \in \mathbb{N} \quad k \cdot u - v \in \mathcal{C}(N)\,. \tag{2}$$

For a proof see § 6.2.1 in [8]. Lemma 1 implies that, given $\langle A, B|C\rangle \in \mathcal{T}(N)$ and $u \in \mathcal{S}(N)$, one has $\langle A, B|C\rangle \in \mathcal{M}_u$ iff $u \rightharpoonup u_{\langle A,B|C\rangle}$. Thus, testing for whether a triplet is represented in u is a special case of testing independence implication.

To illustrate the result let us consider the imset $u = u_{\langle a,bc|\emptyset\rangle}$ over $N = \{a, b, c\}$ shown in Figure 2. By Lemma 1 u i-implies the elementary imset $v = u_{\langle a,b|c\rangle}$. Indeed, for $l = 1$ one has $l\cdot u - v = u_{\langle a,c|\emptyset\rangle} \in \mathcal{E}(N) \subseteq \mathcal{C}(N) \subseteq \mathcal{S}(N)$. Note that in this case $l = 1$ but it may be the case that the constant l must exceed 1 - an example is given in § 6.2.1 of [8].

REMARK 1 Note that there exists an alternative algebraic characterization of independence implication. This dual characterization, named *skeletal characterization* in [8], is based on the existence of a certain special finite collection $\mathcal{K}^{\diamond}_{\ell}(N)$ of non-negative imsets over N, called the *ℓ-skeleton*, such that an imset u over N is structural iff, for every $m \in \mathcal{K}^{\diamond}_{\ell}(N)$,

$$\langle m, u\rangle \equiv \sum_{S\subseteq N} m(S)\cdot u(S) \geq 0\,.$$

Nevertheless, although the existence of $\mathcal{K}^{\diamond}_{\ell}(N)$ was proved (see § 5.2.2 of [8]) its concrete form is only known for $|N| \leq 5$. Moreover, some results suggest that the number of elements of the ℓ-skeleton grows superexponentially with $|N|$ (see § 5.2 in [2] for related discussion). It is shown in § 6.2.2 of [8] that, given $u, v \in \mathcal{S}(N)$, one has $u \rightharpoonup v$ iff, for every $m \in \mathcal{K}^{\diamond}_{\ell}(N)$,

$$\langle m, v\rangle > 0 \quad\Rightarrow\quad \langle m, u\rangle > 0\,. \tag{3}$$

This alternative characterization of independence implication is, therefore, applicable only when the number of variables in a problem is small.

4 Implementation aspects

The aim of this section is to analyze implementation aspects related to testing independence implication of structural imsets. Lemma 1 allows one to transform testing independence implication to a series of tasks of deciding whether an imset is structural. Thus, one needs to analyze this question from a theoretical point of view. The first step is to analyze the same task for combinatorial imsets.

4.1 Testing combinatorial imsets

A basic observation is as follows (for a proof see Proposition 4.3 in [8]).

LEMMA 2 Given $u \in \mathcal{C}(N)$, let

$$\sum_{v\in\mathcal{E}(N)} l_v \cdot v = u = \sum_{v\in\mathcal{E}(N)} k_v \cdot v, \quad l_v, k_v \in \mathbb{Z}^+,$$

be two decompositions of u into elementary imsets. Then $\sum_{v\in\mathcal{E}(N)} l_v = \sum_{v\in\mathcal{E}(N)} k_v$. This number, called the *degree* of u and denoted by $deg(u)$, can be obtained as follows:

$$deg(u) = \langle m_*, u\rangle, \tag{4}$$

where $m_*(S) = \frac{1}{2}\cdot|S|\cdot(|S|-1)$ for every $S \subseteq N$.

Given an imset u over N, the formula (4) gives an integer. If $u \in \mathcal{C}(N)$ then it is non-negative and coincides with the degree of u. Therefore, the complexity of testing whether u is combinatorial ultimately depends on $\langle m_*, u\rangle$, provided it is non-negative. The only combinatorial imset of degree 0 is the zero imset $u \equiv 0$. Moreover, it is clear that an imset u over N is a combinatorial imset of degree n, $n \in \mathbb{N}$ if there exists $u_{\langle a,b|C\rangle} \in \mathcal{E}(N)$ such that $u - u_{\langle a,b|C\rangle}$ is a combinatorial imset of degree $n-1$. In particular, the only combinatorial imsets of degree 1 are elementary imsets which are known. The above fact means that a test for a combinatorial imset of degree n can be transfomed into a number of tests for combinatorial imsets of smaller degree. Therefore, testing combinatorial imsets can be done recursively.

Let me explain that the above observation does not mean that testing combinatorial imsets is an easy task from a practical point of view. It may be time-consuming if the number $\langle m_*, u\rangle$ is very high. What is important is that testing combinatorial imsets can be performed in a finite number of steps and the number of these steps can be estimated in advance! In particular, if the number $\langle m_*, u\rangle$ is low then testing whether $u \in \mathcal{C}(N)$ is quite simple even if $|N|$ is very high.

4.2 Testing structural imsets

It was mentioned in Remark 1 that structural imsets can be recognized by means of a special *skeletal criterion* which is, however, applicable only if the number of variables is at most 5. Therefore, one needs to look for other ways of testing structural imsets. The following lemma allows one to reduce this question to the quesion of testing combinatorial imsets – at least, from a theoretical point of view.

LEMMA 3 There exists a constant $n \in \mathbb{N}$, depending on $|N|$, such that, for every imset u over N,

$$u \in \mathcal{S}(N) \quad \Leftrightarrow \quad n \cdot u \in \mathcal{C}(N)\,. \qquad (5)$$

For a proof see § 6.3.1 in [8] – it uses some special facts from theory of integer programming [3], namely the existence of so-called Hilbert basis of a rational cone.

The above result is a theoretical one; for practical purposes one needs to know the value of the least constant $n \in \mathbb{N}$ satisfying (5). The fact that $\mathcal{S}(N) = \mathcal{C}(N)$ for $|N| \leq 4$ verified in [4] implies that $n = 1$ for $|N| \leq 4$. Since no example of $u \in \mathcal{S}(N) \setminus \mathcal{C}(N)$ is known, it may be the case that $n = 1$ for any $|N|$. Therefore, it is quite important either to verify or to disprove the hypothesis that structural and combinatorial imsets coincide for any $|N|$. If the hypothesis is confirmed then the question of testing structural imsets will be essentially simplified.

4.3 Testing independence implication

It follows from Lemma 1 that $u \rightharpoonup v$ for $u, v \in \mathcal{S}(N)$ iff there exists $l \in \mathbb{N}$ such that $l \cdot u - v \in \mathcal{S}(N)$. A natural question arises if there is an upper limit for the constant l. If $l \cdot u - v \in \mathcal{S}(N)$ then one can show using Lemma 2 and the fact that m_* belongs to the convex cone generated to $\mathcal{K}_\ell^\diamond(N)$ that $l \cdot deg(u) \geq deg(v)$. This indicates that there is no hope for an upper limit of l unless the degree of v is limited. Thus, let us consider the the same question assuming that $deg(v) = 1$, that is, v is an elementary imset. Then one can easily observe the following fact (see § 6.3.2 in [8]).

LEMMA 4 There exists the least constant $l \in \mathbb{N}$, depending on $|N|$, such that, for every $u \in \mathcal{S}(N)$ and $v \in \mathcal{E}(N)$,

$$u \rightharpoonup v \quad \Leftrightarrow \quad l \cdot u - v \in \mathcal{S}(N)\,. \qquad (6)$$

For practical purposes one needs to know what is the actual value of the least constant l. It is known for $|N| \leq 5$: one has $l = 1$ if $|N| \leq 4$ and $l = 7$ for $|N| = 5$ (this can be shown with the aid of an example in § 4.3 of [6]). Actually, I have a hypothesis that the least constant l can be obtained on basis of the ℓ-skeleton (see Remark 1) as the following number, called the *grade* and denoted by $gra(N)$:

$$\max\{\, \langle m, v\rangle\,;\; m \in \mathcal{K}_\ell^\diamond(N),\; v \in \mathcal{E}(N)\,\}\,.$$

The hypothesis was verified for $|N| \leq 5$. Another supportive argument for the hypothesis is a theoretical result (Lemma 6.4 in [8]) which allows one to express, on basis of the ℓ-skeleton, the minimal constant l such that (6) is true for every $u \in \mathcal{C}(N)$ and $v \in \mathcal{E}(N)$. Actually, if $\mathcal{S}(N) = \mathcal{C}(N)$ and another hypothesis saying that, for every $m \in \mathcal{K}_\ell^\diamond(N)$,

$$\min \{ \langle m, w \rangle ;\ w \in \mathcal{E}(N),\ \langle m, w \rangle \neq 0 \} = 1,$$

is true then the above mentioned result implies that the minimal constant l satisfying the condition in Lemma 4 is just the grade *gra*(N).

5 Linear programming approach

One of the reviewers draw my attention to another method for testing independence implication (in addition to those mentioned in § 4.3 and in Remark 1). The basic idea is to transform the implication task to a classic maximization problem of linear programming. This idea of Yeung was already mentioned by Matúš in the end of § 5 of [2] in slightly different but analogous framework and later elaborated in § 13.2–13.5 of [9] as a tool for deriving information-theoretical inequalities.

To explain the idea in the framework of this paper consider the class $\mathcal{K}_\ell(N)$ of ℓ-standardized supermodular functions on the power set of N, that is, the class of functions $m : \mathcal{P}(N) \to \mathbb{R}$ such that

$$\begin{array}{ll} m(S) = 0 & \text{whenever } |S| \leq 1\,, \\ \langle m, w \rangle \geq 0 & \text{for every } w \in \mathcal{E}(N). \end{array}$$

As $m \geq 0$ for every $m \in \mathcal{K}_\ell(N)$ the class $\mathcal{K}_\ell(N)$ is a pointed rational cone. Note that the elements of the ℓ-skeleton $\mathcal{K}_\ell^\diamond(N)$ mentioned in Remark 1 correspond to extreme rays of $\mathcal{K}_\ell(N)$. One can show (see Lemma 6.2 in [8]) that, given $u, v \in \mathcal{S}(N)$, one has $u \rightharpoonup v$ iff

$$\langle m, u \rangle = 0 \ \Rightarrow\ \langle m, v \rangle = 0$$

for every $m \in \mathcal{K}_\ell(N)$. Therefore, one can consider the maximization problem

$$\max \{ \langle m, v \rangle ;\ m \in \mathcal{K}_\ell(N)\ \langle m, u \rangle = 0 \} \quad (7)$$

and observe that the maximum in (7) is 0 iff u i-mplies v; note that the function to be maximized is non-negative on the respective domain. The reviewer thinks that one can compute the maximum in (7) by means of standard software packages (based on the simplex method) and this is feasible beyond the case $|N| \leq 5$. I am not as optimistic as the reviewer is. Let me give two reasons why I am more sceptical about the effectiveness of this approach.

First, I have checked how the proposed method works in a very simple special case $N = \{a, b, c\}$, $u = u_{\langle a,bc|\emptyset \rangle}$ and $v_1 = u_{\langle a,c|\emptyset \rangle}$, respectively $v_2 = u_{\langle b,c|\emptyset \rangle}$. I have used the simplex method with a known vertex of the respective domain described in § 11.1 of [3]. While the methods described in Remark 1 and § 4.3 give an immediate result in these two cases the application of the simplex method requires a lot of computation. To give the reader some idea let me explain that the considered simplex method is iterative and, if applied to (7), in each iteration a plenty of arithmetic operations are done. More specifically, in each iteration, a system of linear equations $\mathbf{A} \cdot \mathbf{x} = \mathbf{y}$ is solved twice, where $\mathbf{A}$ is a regular matrix of the rank $2^{|N|} - |N| - 1$, scalar products $\langle m, w \rangle$, $w \in \mathcal{E}(N)$ are computed for certain $m : \mathcal{P}(N) \to \mathbb{R}$, some coefficients are computed and their minimum is found. In the considered example, one needs 3 iterations to derive $u \rightharpoonup v_1$ and 4 iterations to show $\neg(u \rightharpoonup v_2)$. This is quite complicated way of verifying it and one can hardly believe that the same method will work effectively if the number of variables increases.

The second argument for my scepticism is intuitive. However, because reviewer's opinion is also based on his intuitive belief rather than on a real experiment, I think that relevant intuitive counter-arguments should also be mentioned. The underlying idea of the classic simplex method is that one moves from one vertex of a polytope to another one untill an optimal vertex is reached. In the context of (7) it corresponds to the effort to find an optimal extreme ray of the cone which serves as the domain in (7). Note that because the domain is a face of $\mathcal{K}_\ell(N)$ this optimal ray is also an extreme ray of $\mathcal{K}_\ell(N)$. To that end several tentative moves from the original zero vector $m_0 \equiv 0$ are generated and tested for feasibility – the resulting vector m must belong to the domain of (7). Every iteration of the simplex method corresponds to a tentative move of this kind. The argument to show that the maximum in (7) is greater that 0, that is, to disprove $u \rightharpoonup v$, is that an element $m \in \mathcal{K}_\ell^\diamond(N)$ is found such that $\langle m, u \rangle = 0$ and $\langle m, v \rangle > 0$. However, this is the same argument which is behind the method mentioned in Remark 1!

Using the results of Chapter 5 of [8] I can show that for every element $m' \in \mathcal{K}_\ell^\diamond(N)$ of the ℓ-skeleton, there exists $u, v \in \mathcal{S}(N)$ such that the only elements of $\mathcal{K}_\ell(N)$ which can be used to disprove $u \rightharpoonup v$ are non-zero multiples of m'. That means, in order to be able to disprove the independence implication the simplex method has to be able to find every element of the ℓ-skeleton. From this point of view, the main difference between the simplex method and the method mentioned in Remark 1 is that, in the former case, a suitable element of $\mathcal{K}_\ell^\diamond(N)$ is searched following a certain blind scenario, while in the latter case all elements of $\mathcal{K}_\ell^\diamond(N)$ were obtained in advance and one only needs to compute scalar products of u and v with them. Thus, the disadvantage of the simplex method seems to be that time-consuming computation of extreme rays of $\mathcal{K}_\ell(N)$ is repeated during the testing procedure while its advantage seems to be that the fixing imsets u and v may perhaps avoid moves towards 'wrong' extreme rays. I am not an expert in the simplex method for which reason I can hardly estimate which of these two phenomena will appear to be more influential. Despite my scepticism I agree that the linear programming approach deserves a comparison with the other two methods of testing independence implication; for example, by a thorough computer experiment.

To conclude my comment let me explain that the aim of this paper is to explore a theoretical basis for computer testing of independence implication. Because I am a mathematician, my primary goal is deeper understanding of the problem from the theoretical point of view rather than the immediate use of any available method. The intution obtained as a result of a theoretical analysis can hopefully be utilized to propose an effective method for testing independence implication. I hope that a long-term research effort in this direction will result in a more effective method than the blind use of such kind of an algorithm for which it is not known how much time is needed to get an answer.

6 Discussion

The reader may find it surprising, but the method of structural imsets can even find some applications in the area of graphical models. I think that two of them deserve mentioning.

1. One of common problems in the area of Bayesian networks is to recognize independence equivalent acyclic directed graphs, that is, graphs describing the same CI structure. The method of structural imsets offers, through the concept of a *standard imset*, a fairly good algebraic criterion to recognize this equivalence of acyclic directed graphs – see Corollary 7.1 in [8]. Actually, there is also an algebraic criterion for testing inclusion of CI structures described by those graphs (Lemma 8.6 of [8]).

2. An algebraic point of view brings a new simplifying perspective to some existing methods of learning Bayesian networks, namely those based on the maximization of a suitable quality criterion – for details see Chapter 8 of [8]. The point is that usual criteria used in practice appear to be affine functions of the respective standard imsets. It seems that this point of view can be extended and can lead to a proposal of a general method for learning CI structures.

Of course, there are many interesting open problems related to the method of structural imsets, both theoretical and practical ones – see Chapter 9 of [8] for their overview. They will be a topic of a future research.

Acknowledgements

I am indebted to my colleague František Matúš and to both reviewers for their comments.

References

[1] S. L. Lauritzen (1996). Graphical Models. *Clarendon Press.*

[2] F. Matúš (1997). Conditional independence structures examined via minors. *Annals of Mathematics and Artificial Intelligence 21*, pp. 99-128.

[3] A. Schrijver (1986). Theory of Linear and Integer Programming. *John Wiley.*

[4] M. Studený (1991). Convex set functions I. and II. Research reports n. 1733 and n. 1734, Institute of Information Theory and Automation, Prague, November 1991.

[5] M. Studený (1994/95). Description of structures of stochastic conditional independence by means of faces and imsets (a series of 3 papers). *International Journal of General Systems 23*, pp. 123-137, 201-219, 322-341.

[6] M. Studený, R. R. Bouckaert, T. Kočka (2000). Extreme supermodular set functions over five variables. Research report n. 1977, Institute of Information and Automation, Prague, January 2000.

[7] M. Studený (2001). On non-graphical description of models of conditional independence structure. A paper presented at *HSSS Workshop on Stochastic Systems for Individual Behaviours, January 22-24, 2001, Louvain-la-Neuve, Belgium*, available at http://www.utia.cas.cz/user_data/studeny/ff16.html

[8] M. Studený (2004). On Probabilistic Conditional Independence Structures. To be published by *Springer-Verlag.*

[9] R. W. Yeung (2002). A First Course in Information Theory. *Kluwer.*