On Open Questions in the Geometric Approach to Learning BN Structures

Milan Studený
Dept. of Decision-Making Theory
ÚTIA AV ČR, v.v.i.
studeny@utia.cas.cz

Jiří Vomlel
Dept. of Decision-Making Theory
ÚTIA AV ČR, v.v.i.
vomlel@utia.cas.cz

Abstract

The basic idea of an algebraic approach to learning Bayesian network (BN) structures is to represent every BN structure by a certain (uniquely determined) vector, called the *standard imset*. In a recent paper [11], we have shown that the set S of standard imsets is the set of vertices (= extreme points) of a certain polytope P and introduced natural *geometric neighborhood* for standard imsets, and, consequently, for BN structures.

The new geometric view led to a series of open mathematical questions. In this contribution, we try to answer some of them. First, we introduce a class of necessary linear constraints on standard imsets and formulate a conjecture that these constraints characterize the polytope P. The conjecture has been confirmed in the case of (at most) 4 variables. Second, we confirm a former hypothesis by Raymond Hemmecke that the only lattice points (= vectors having integers as components) within P are standard imsets. Third, we give a partial analysis of the geometric neighborhood in the case of 4 variables.

1 Motivation

The motivation for this research is learning Bayesian network (BN) structures from data by the method of maximization of a quality criterion (= the score and search method). By a *quality criterion* is meant a real function $\mathcal{Q}$ of the BN structure (= of a graph G, usually) and of the database D. The value $\mathcal{Q}(G, D)$ should say how much the BN structure given by G is good to explain the occurrence of the database D.

The basic idea of an algebraic and geometric approach to this topic, proposed in Chapter 8 of [8] and then developed in [11], is to represent the BN structure given by an acyclic directed graph G by a certain vector u_G having integers as components, called the *standard imset* (for G). The point is that then every reasonable criterion $\mathcal{Q}$ for learning BN structures (score equivalent and decomposable one) is an affine function (= a linear function plus a constant) of the standard imset. More specifically, one has

$$\mathcal{Q}(G,D) \,=\, s_D^{\mathcal{Q}} - \langle t_D^{\mathcal{Q}}, u_G\rangle,$$

where $s_D^{\mathcal{Q}}$ is a real number, $t_D^{\mathcal{Q}}$ a vector of the same dimension as the standard imset u_G (these parameters both depend solely on the database D and the criterion $\mathcal{Q}$) and $\langle *, * \rangle$ denotes the scalar product. The vector $t_D^{\mathcal{Q}}$ is named the *data vector* (relative to $\mathcal{Q}$).

The main result of [11] is that the set of standard imsets over a fixed set of variables N is the set of vertices (= extreme points) of a certain polytope P. Thus, as every reasonable quality criterion $\mathcal{Q}$ can be viewed as (the restriction of) an affine function on the respective Euclidean space (of higher dimension), the task to maximize $\mathcal{Q}$ over BN structures is equivalent to the task to maximize an affine function over the above-mentioned polytope P.

This maximization problem has been treated thoroughly within the linear programming community. A classic tool to solve linear programming problems is the *simplex method* [5]. One of possible interpretations of this method is that it is a kind of a search method, in which one moves between the vertices of the polytope along its edges (in the geometric sense) until an optimal vertex is reached. This motivated the concept of the *geometric neighborhood* for standard imsets, and, consequently, for BN structures.

Several open mathematical questions have been mentioned in the conclusions of [11]. They are motivated by the above-mentioned intention to apply linear programming methods in the area of learning BN structures. This contribution is devoted to three of them.

2 Basic concepts

2.1 Learning BN structures

Throughout this paper we assume that N is a non-empty finite set of *variables*. Every variable $i \in N$ is assigned a finite set of possible values, the individual sample space X_i. To avoid trivial cases and consequent troubles we assume $|\mathsf{X}_i| \geq 2$ for any $i \in N$.

Let $\mathsf{DAGS}(N)$ denote the collection of all acyclic directed graphs having N as the set of nodes. The (discrete) *Bayesian network* (BN) is a pair (G,P), where $G \in \mathsf{DAGS}(N)$ and P is a probability distribution on the joint sample space $\mathsf{X}_N \equiv \prod_{i\in N} \mathsf{X}_i$ which (recursively) factorizes according to G [4]. Given $G \in \mathsf{DAGS}(N)$, the respective statistical model of a *BN structure* is the class of all distributions P on X_N that factorize according to G.

Note it may happen that two different graphs over N describe the same BN structure. Thus, one is usually interested in describing the BN structure by a unique representative. A classic such graphical representative is a special chain graph, called the *essential graph* [1]. However, in our algebraic approach, we use an algebraic representative instead, called the *standard imset* (see below).

There is a polynomial algorithm for transforming the essential graph into the standard imset and conversely [10].

Learning BN structures is done on the basis of data, assumed in the form of a complete database $D : x^1, \ldots, x^d$ of the length $d \geq 1$, which is a sequence of elements of the joint sample space X_N. Let $\mathsf{DATA}(N, d)$ denote the collection of all databases from X_N of the length d. A *quality criterion* (for learning BN structures) is a real function $\mathcal{Q}$ on $\mathsf{DAGS}(N) \times \mathsf{DATA}(N, d)$. The learning procedure based on $\mathcal{Q}$ consists in maximizing the function $G \mapsto \mathcal{Q}(G, D)$ over $G \in \mathsf{DAGS}(N)$, where $D \in \mathsf{DATA}(N, d)$ is the observed database. Thus, the value $\mathcal{Q}(G, D)$ should somehow evaluate how the statistical model determined by G fits the database D. We refer for the related concept of (statistical) consistency of a quality criterion to § 8.4.2 in [4].

However, there are other technical requirements on quality criteria raised in connection with computational methods for their maximization [3]. A criterion is *decomposable* if it is the sum of contributions that correspond to factors in the factorization according to the graph and *score equivalent* [2] if it ascribes the same value to graphs describing the same BN structure. There are several examples of quality criteria that meet these requirements. A kind of standard example of such a criterion is Schwarz's *Bayesian information criterion* (BIC) [6], but there is also a bunch of Bayesian quality criteria [9].

2.2 A few concepts from polyhedral geometry

Let us consider a real Euclidean space $\mathbb{R}^K$, where K is a non-empty finite set. The scalar product of two vectors $\boldsymbol{v}, \boldsymbol{x}$ in $\mathbb{R}^K$ will be denoted as follows:

$$\langle \boldsymbol{v}, \boldsymbol{x} \rangle \equiv \sum_{s \in K} v_s \cdot x_s \, .$$

A *rational polytope* in $\mathbb{R}^K$ is the convex hull of a finite set $\mathsf{V} \subseteq \mathbb{Q}^K$ of rational points. A well-known result in polyhedral geometry (Corollary 7.1c in [5]) says that a polytope can equivalently be characterized by means of a finite number of linear inequality constraints.

Note that the classic version of the *simplex method* is applicable to the task to find maximum/minimum of a linear function over a set $\mathsf{P} \subseteq \mathbb{R}^K$ defined by means of a finite number of linear inequality constraints (see Chapter 11 in [5]).

A *vertex* (= an extreme point) of a polytope P is a point in P which cannot be written as a convex combination of other elements in P. An *edge* of P is a line-segment $[\boldsymbol{x}, \boldsymbol{y}]$, where $\boldsymbol{x}, \boldsymbol{y}$ are distinct vertices of the polytope P and the set $\mathsf{P} \setminus [\boldsymbol{u}, \boldsymbol{v}]$ is convex. The vertices and edges of a polytope are quite important in linear programming because the simplex method applied to a polytope P can be interpreted as a kind of search method in which one moves between the vertices of P along its (geometric) edges (see § 11.1 of [5]).

2.3 Imsets

The method of *structural imsets* has been proposed in [8] to provide an universal (mathematical) tool for describing probabilistic conditional independence structures. In the context of graphical models, it leads to an algebraic approach to learning BN structures.

An *imset* u over N is an integer-valued function on $\mathcal{P}(N) \equiv \{A;\, A \subseteq N\}$, the power set of N. It can be viewed as a vector whose components are integers, indexed by subsets of N. Any real function $m : \mathcal{P}(N) \to \mathbb{R}$ will be analogously interpreted as a real vector (= identified with an element of $\mathbb{R}^{\mathcal{P}(N)}$). Thus, an imset is nothing but an element of $\mathbb{Z}^{\mathcal{P}(N)}$; in the context of integer programming [5] called a *lattice point* in the Euclidean space $\mathbb{R}^{\mathcal{P}(N)}$.

A trivial example of an imset is the *zero imset*, denoted by 0. Given $A \subseteq N$, the symbol δ_A will denote this basic imset:

$$\delta_A(B) = \begin{cases} 1 & \text{if } B = A, \\ 0 & \text{if } B \neq A, \end{cases} \qquad \text{for } \; B \subseteq N.$$

Since $\{\delta_A;\, A \subseteq N\}$ is a linear basis of $\mathbb{R}^{\mathcal{P}(N)}$, any imset can be expressed as a combination of these basic imsets.

An *elementary imset* (over N) is an imset of the form

$$u_{\langle a,b|C\rangle} \;=\; \delta_{\{a,b\}\cup C} + \delta_C - \delta_{\{a\}\cup C} - \delta_{\{b\}\cup C}\,,$$

where $C \subseteq N$ and $a, b \in N \setminus C$ are distinct. In our algebraic approach [8] it encodes an elementary conditional independence statement $a \perp\!\!\!\perp b \,|\, C$. The class of all elementary imsets over N will be denoted by $\mathcal{E}(N)$; it is a finite subset of $\mathbb{R}^{\mathcal{P}(N)}$. The cone spanned by $\mathcal{E}(N)$ will be denoted by $\mathcal{R}(N)$.[1]

An imset will be called *combinatorial* if it is a combination of elementary imsets with non-negative integers as coefficients.[2] The *degree* of a combinatorial imset u, denoted by $deg\,(u)$, is the number

$$deg\,(u) = \langle m_*, u\rangle \equiv \sum_{S \subseteq N} m_*(S) \cdot u(S)\,, \tag{1}$$

where $m_*(S) = \frac{1}{2} \cdot |S| \cdot (|S| - 1)$ for $S \subseteq N$. It is shown in Proposition 4.3 of [8] that $deg\,(u)$ is the sum of coefficients in the decomposition of u into elementary imsets; in particular, this sum only depends on u, not on a particular combination of elementary imsets yielding u.

2.4 Algebraic approach to learning BN structures

Given $G \in \mathsf{DAGS}(N)$, the *standard imset* for G is given by the formula:

$$u_G \;=\; \delta_N - \delta_\emptyset + \sum_{i \in N} \{\, \delta_{pa_G(i)} - \delta_{\{i\} \cup pa_G(i)} \,\}, \tag{2}$$

[1] It is a pointed rational polyhedral cone in $\mathbb{R}^{\mathcal{P}(N)}$.

[2] Equivalently, the sum of elementary imsets with allowed repetition of summands.

where $pa_G(i) = \{\, j \in N;\ j \to i \ \text{ in } G \,\}$ denotes the set of *parents* if i in G. Note that the terms in (2) can both sum up and cancel each other. Nevertheless, it follows from the definition that u_G has at most $2 \cdot |N|$ non-zero values. Thus, the memory demand for representing standard imsets are polynomial in $|N|$.

An important observation is that, for $G, H \in \mathsf{DAGS}(N)$, one has $u_G = u_H$ iff they describe the same BN structure (Corollary 7.1 in [8]). In particular, the standard imset for $G \in \mathsf{DAGS}(N)$ is a unique representative of the corresponding BN structure. Note that every standard imset is combinatorial; actually, it is a sum of elementary imsets (see Lemma 2 in § 5). The degree of a standard imset u_G is $\binom{|N|}{2} - r$, where r is the number of arrows in G (see Lemma 7.1 in [8]).

Now, Lemmas 8.3 and 8.7 from [8] together say that every score equivalent and decomposable criterion $\mathcal{Q}$ must have the form:

$$\mathcal{Q}(G, D) \;=\; s^{\mathcal{Q}}_D - \langle t^{\mathcal{Q}}_D, u_G \rangle \quad \text{for } G \in \mathsf{DAGS}(N), D \in \mathsf{DATA}(N, d), d \geq 1 \qquad (3)$$

where the constant $s^{\mathcal{Q}}_D \in \mathbb{R}$ and the vector $t^{\mathcal{Q}}_D : \mathcal{P}(N) \to \mathbb{R}$ do not depend on G. The formulas for the data vector $t^{\mathcal{Q}}_D$ relative to some basic quality criteria $\mathcal{Q}$ have been derived in [8, 9].

2.5 Geometric view on learning BN structures

Let us take a geometric view on the set of standard imsets over a fixed set of variables N, denoted by S:

$$\mathsf{S} \equiv \{\, u_G;\ G \in \mathsf{DAGS}(N) \,\} \subseteq \mathbb{R}^{\mathcal{P}(N)}\,.$$ [3]

Theorem 4 in [11] says that S is the set of vertices of a rational polytope $\mathsf{P} \subseteq \mathbb{R}^{\mathcal{P}(N)}$, whose dimension is $2^{|N|} - |N| - 1$. This polytope P will be called the *standard imset polytope* in the sequel. It follows from (3) that the task to maximize $\mathcal{Q}$ over $G \in \mathsf{DAGS}(N)$ is equivalent to the task to minimize the linear function $u \mapsto \langle t^{\mathcal{Q}}_D, u \rangle$ over P.

The idea of application of linear programming methods in the area of learning BN structures led to the concept of geometric neighborhood for BN structures. More specifically, two standard imsets $u, v \in \mathsf{S}$ will be called the *geometric neighbors* if the line-segment connecting them in $\mathbb{R}^{\mathcal{P}(N)}$ is an edge of the standard imset polytope P.

It has been shown in Theorem 5 of [11] that the well-known *inclusion neighborhood*, used widely in present computational methods for learning BN structures, like the GES algorithm [3], is strictly contained in the geometric one. Moreover, it follows from Corollary 8.4 in [8] that standard imsets $u, v \in \mathsf{S}$ correspond to inclusion neighbors iff their differential imset $w = u - v$ is either elementary one or a multiple of it by -1.

The importance of the concept of geometric neighborhood is based on the fact that, for any affine function $\mathcal{Q}$ on P, a local maximum of $\mathcal{Q}$ in $u \in \mathsf{S}$ with

[3] To avoid misunderstanding recall that distinct $G, H \in \mathsf{DAGS}(N)$ may give rise the same standard imset $u_G = u_H$ but S contains just one vector for any group of graphs defining the same BN structure.

respect to the geometric neighborhood must be the global maximum of $\mathcal{Q}$ over P (Theorem 6 in [11]). In particular, this holds for any reasonable quality criterion $\mathcal{Q}$ for learning BN structures. The following research goals have been expressed in conclusions of [11]:

- Describe linear constraints on the elements P. A complete characterization of these constraints would provide a description of P suitable for the intended application of linear programming methods.
- An interesting related conjecture by Raymond Hemmecke is that the only lattice points within P are standard imsets.
- Describe *differential imsets* for geometric neighbors, that is, imsets of the form $u_G - u_H$, where $G, H \in \mathsf{DAGS}(N)$ are such that u_G and u_H are geometric neighbors.

These questions concern the complexity of a potential future linear programming procedure for maximization of a quality criterion $\mathcal{Q}$. In this paper we answer partially some of them.

3 Necessary linear constraints

In this section, we summarize all linear constraints on standard imsets we are aware of. Of course, they give necessary conditions on points in P.

3.1 Overview of the constraints

We classify our linear constraints into three groups, denoted (A), (B) and (C). First, basic results from [8] imply that every standard imset belongs to the cone $\mathcal{R}(N)$ generated by elementary imsets. This observation implies two kinds of necessary linear conditions on the elements of P: the equality constraints, denoted by (A), and the remaining inequality constraints, denoted by (B).

(A) Equality constraints

If $u \in \mathsf{S}$ then the following two conditions are valid:

$$(A.1) \qquad \sum_{S,\, S\subseteq N} u(S) = 0,$$

$$(A.2) \qquad \forall\, i \in N \qquad \sum_{S,\, i\in S\subseteq N} u(S) = 0.$$

This means that S, and, therefore, P as well, belongs to a linear subspace of $\mathbb{R}^{\mathcal{P}(N)}$ of the dimension $2^{|N|} - |N| - 1$.

Table 1: Number of non-specific inequality constraints.

$\lvert N\rvert$	2	3	4	5
$\lvert\mathcal{K}_\ell^\diamond(N)\rvert$	1	5	37	117978

(B) Non-specific inequality constraints

The inequality constraints for points in the cone $\mathcal{R}(N)$ are related to supermodular functions. A function $m : \mathcal{P}(N) \to \mathbb{R}$ is called *supermodular* iff

$$m(C \cup D) + m(C \cap D) \geq m(C) + m(D) \qquad \text{for every } C, D \subseteq N.$$

An equivalent definition is that $\langle m, v\rangle \geq 0$ for every elementary imset v over N. This observation gives a (formally infinite) set of inequality constraints for the points in $\mathcal{R}(N)$, and, therefore, for any standard imset u:

$$(B) \qquad \langle m, u\rangle \geq 0 \quad \text{for every supermodular function } m : \mathcal{P}(N) \to \mathbb{R}.$$

Nevertheless, the point is that this condition can equivalently be formulated in the form of a finite number of linear inequality constraints. First, without loss of generality one can assume that $m(S) = 0$ for $S \subseteq N$ with $|S| \leq 2$. Second, the class of these special supermodular functions is a pointed rational polyhedral cone and has, therefore, finitely many extreme rays.[4] Thus, the class normalized integral representatives of these extreme rays, denoted by $\mathcal{K}_\ell^\diamond(N)$ and called the ℓ-skeleton in [8], establishes a finite set of normalized inequality constraints:

$$\forall\, m \in \mathcal{K}_\ell^\diamond(N) \qquad \langle m, u\rangle \geq 0.$$

These (representatives of) extreme rays have been computed for $|N| \leq 5$ using linear programming packages [7]. It seems that the number of these extreme rays grows super-exponentially with $|N|$; their numbers are in Table 1.

It looks like none of these inequality constraints for P is derivable from the other constraints (including those mentioned below).

(C) Specific inequality constraints

The results of [10] led to a series of specific linear inequality constraints for standard imsets, that are not valid for all points in the cone $\mathcal{R}(N)$. These constraints are related to "ascending" classes of sets. We say that a class $\mathcal{A} \subseteq \mathcal{P}(N)$ of subsets of N is *closed under supersets* if

$$\forall\, S \in \mathcal{A} \quad \text{if } S \subseteq T \subseteq N \quad \text{then } T \in \mathcal{A}.$$

To avoid vacuous constraints and a trivial consequence of (A.1) we consider only non-empty classes of non-empty sets. This gives the following series of

[4] See § 5.1.2 and Lemma 5.3 (pp. 90-93) in [8] for both these claims.

constraints:

$$(C) \qquad \sum_{S\in\mathcal{A}} u(S) \leq 1 \quad \text{for any system } \emptyset \neq \mathcal{A} \subseteq \{S \subseteq N; |S| \geq 1\} \text{ which is closed under supersets.}$$

Note that, unlike the number (B)-constraints, the number of constraints in (C) seems to grow only exponentially with $|N|$. Actually, these constraints are in correspondence with log-linear models over N.[5] Nevertheless, the list of conditions (C) is not reduced completely: some of these constraints are superfluous because they follow from the other ones combined with (A) and (B).[6] Moreover, each of the (C)-constraints can, owing to (A.2), be re-formulated equivalently in a kind of "standard" form

$$\sum_{S\in\mathcal{B}} k_S \cdot u(S) \leq 1 \quad \text{for } \mathcal{B} \subseteq \{S \subseteq N;\, |S| \geq 2\} \text{ and } k_S \in \mathbb{Z} \text{ for } S \in \mathcal{B}.$$

It looks like none of the constraints for $\mathcal{A} \subseteq \{S \subseteq N;\, |S| \geq 2\}$ is superfluous, while if $\mathcal{A}$ contains a singleton then both cases can occur: the respective inequality constraints can be either superfluous or non-derivable from others.[7]

Lemma 1. (the necessity of specific constraints)
If $u \in \mathsf{S}$ is a standard imset over N then the condition (C) is valid.

The proof is omitted because of limited scope of a conference contribution.

3.2 Conjecture about the linear constraints

The constraints (A)-(C) from the preceding section have several consequences, which are, perhaps, not evident at first sight. One of them is that every standard imset $u \in \mathsf{S}$ is bounded from below: $u(S) \geq -1$ for any $S \subseteq N$.

We have shown that (A)-(C) are necessary constraints on points in P, but we have also some reasons to conjecture that they are sufficient to characterize the standard imset polytope P. More specifically, we have verified for $|N| \leq 4$ that the conditions (A)-(C) characterize P. Thus, we dare to formulate the following hypothesis.

> *Conjecture* The linear constraints (A)-(C) together form a necessary and sufficient condition for $u \in \mathbb{R}^{\mathcal{P}(N)}$ to belong to P.

---

[5]This is because every class of sets closed under supersets is determined by the collection of its minimal sets, which is a class of incomparable sets. Hierarchical log-linear models also correspond to classes of incomparable subsets of the class $\{A \subseteq N; |S| \geq 1\}$, namely to those whose union is N.

[6]For example, if $\mathcal{A} = \{S \subseteq N;\ i \in S\}$ for some $i \in N$ then (A.2) gives $\sum_{S\in\mathcal{A}} u(S) = 0 \leq 1$.

[7]This happens in the case $|N| = 5$.

4 Lattice points in the standard imset polytope

Another related question concerning the polytope P is how "thick" it is. More specifically, we may ask whether there exists a lattice point in its interior. Raymond Hemmecke made some computations to find out whether such a point exists in the case $|N| \leq 5$ and the result was negative. This led him to a hypothesis that every lattice point in the standard imset polytope is already the standard imset. In this paper, we confirm the hypothesis:

Theorem 1. *If* $u \in \mathsf{P} \cap \mathbb{Z}^{\mathcal{P}(N)}$ *then* $u \in \mathsf{S}$.

The proof is quite technical and strongly depends on former results of ours [10]; specifically, it depends on the details of an algorithm for testing whether an imset is standard. It is omitted in this contribution.

In light of Theorem 1 one can formulate a weaker version of the conjecture from § 3.2:

> *Conjecture** The constraints (A)-(C) together form a necessary and sufficient condition for $u \in \mathbb{Z}^{\mathcal{P}(N)}$ to be a standard imset (over N).

Indeed, if *Conjecture* is true then, by Theorem 1, *Conjecture** holds as well. However, it is not clear at this moment whether the proof of *Conjecture** is enough to confirm the hypothesis from § 3.2.

5 Differential imsets over 4 variables

The result of our analysis of the geometric neighborhood in the case $|N| = 4$ is an electronic catalogue. To describe the catalogue we need a few auxiliary observations.

5.1 Some auxiliary concepts and results

Given a differential imset $w = u - v$ for $u, v \in \mathsf{S}$ it follows from the formula (1) that the *degree difference* $deg\,(u) - deg\,(v)$ does not depend on the choice of the pair $u, v \in \mathsf{S}$. This seems to be quite important characteristic of w.

We say that two imsets u, v over N are *permutation equivalent* if there exists a bijection $\pi : N \to N$ such that, for all $S \subseteq N$, it holds that $u(S) = v(\pi(S))$, where $\pi(S) = \{\pi(i);\, i \in S\}$. Each class of permutation equivalent imsets will be called a *PE class*. From the point of view of our analysis, it is not necessary to distinguish between permutation equivalent differential imsets. Every PE class can be described by an arbitrary representative.

Evidently, if $w = u - v$ is a differential imset for $u, v \in \mathsf{S}$ then $-w = v - u$ is a differential imset, too. Again, from the point of view of our analysis it is not necessary to distinguish between w and $-w$. Therefore, we keep only one of these in the catalogue. If the degree difference is non-zero we choose $w = u - v$ with $deg\,(u) > deg\,(v)$. That means, our catalogue only contains (PE representatives of) differential imsets with non-negative degree difference.

An important question is how to express the differential imsets. An elegant solution is offered below.

Lemma 2. *Every standard imset is a combination of elementary imsets with coefficients* $+1$ *(and* 0*).*

A kind of consequence of Lemma 2 is the following observation.

Lemma 3. *Every differential imset* $w = u - v$ *for* $u, v \in \mathsf{S}$ *is a combination of elementary imsets with coefficients* $+1$ *and* -1 *(and* 0*). Moreover, there exists a combination with at most* $\binom{|N|}{2}$ *non-zero coefficients.*

Proofs are omitted because of limited scope of the contribution. In particular, every differential imset for a pair of geometric neighbors can be expressed in that way, which we utilize in our catalogue.

5.2 Description of the catalogue

Our catalogue contains differential imsets $w = u - v$ for those $u, v \in \mathsf{S}$ that are *geometric neighbors*. It contains just one representative for each PE class and only imsets with non-negative degree difference are kept there.

We classified those differential imsets w using three criteria:

- the degree difference for w,
- the *squared Euclidean length* of w, that is, $\sum_{S \subseteq N} w(S)^2$, and
- the number of non-zero imset values, that is, $|\{S \subseteq N;\, w(S) \neq 0\}|$.

In the case $|N| = 4$ the degrees of differential imsets (for geometric neighbors) are integers from the interval $[0, 3]$. The values of the squared Euclidean length are even numbers from interval $[4, 22]$. The numbers of non-zero imset values are integers from interval $[4, 12]$.

There are 8518 ordered pairs (u, v) of geometric neighbors. As explained above, for each couple of ordered pairs (u, v) and (v, u), we have chosen only one differential imset out of $w = u - v$ and $-w = v - u$. In this way, we got 2831 differential imsets; they constitute 319 PE classes. Table 2 gives these numbers for each degree difference.

In order to understand better the geometric neighborhood we searched for an elegant description of differential imsets. One possible solution is offered by Lemma 3: every differential imset over 4 variables can be written as a combination (with coeficients $+1$ or -1) of at most 6 elementary imsets (out of 24 possible elementary imsets).

A complete catalogue of differential imsets over 4 variables with a detailed analysis for each differential imset is available at:

`http://staff.utia.cas.cz/vomlel/imset/catalogue-diff-imsets-4v.html`

Table 2: Numbers of geometric neighbor pairs, differential imsets and PE classes.

degree difference	neigh. pairs	diff. imsets	PE classes
0	2894	927	88
1	4248	1359	144
2	1296	505	71
3	80	40	16
total	8518	2831	319

5.3 A simple example

As mentioned in § 2.5, the classic inclusion neighborhood is contained in the geometric one and the inclusion neighbors are geometric neighbors with the degree difference ± 1.

One of our previous open questions was whether the converse holds. However, as one can deduce from Table 2, this is not true for $|N| = 4$: there are 144 PE classes with the degree difference 1 while one has only 3 PE classes of elementary imsets.

A simple example of a differential imset $w = u - v$ for geometric neighbors $u, v \in \mathsf{S}$ with degree difference 1 that is not an elementary imset is as follows:

$$w = \delta_{\{a\}} - \delta_{\{a,b\}} - \delta_{\{c,d\}} + \delta_{\{b,c,d\}},$$

where

$$u = \delta_{\emptyset} - \delta_{\{a,b\}} - \delta_{\{c,d\}} + \delta_{\{a,b,c,d\}}, \quad v = \delta_{\emptyset} - \delta_{\{a\}} - \delta_{\{b,c,d\}} + \delta_{\{a,b,c,d\}}.$$

6 Conclusions

Let us mention some of our research goals motivated by the results reported here. First, we would like either confirm or disprove the conjecture from § 3.2 for $|N| = 5$. If it is confirmed for $|N| = 5$ we may try to verify the weaker version of the conjecture from § 4 then.

The catalogue from § 5 is meant as a step towards a deeper analysis of the geometric neighborhood. For example, we would like to find out whether there is a graphical interpretation of geometric neighbors, namely whether differential imsets (for geometric neighbors) correspond to graphical operations with the respective essential graphs.

Acknowledgements

This research has been supported by the grants GAČR n. 201/08/0539 and MŠMT n. 1M0572. J. Vomlel has also been supported by the grants GAČR n. 2C06019 and Eurocores LogICCC n. ICC/08/E010.

References

[1] Andersson S.A., Madigan D., Perlman M.D. (1997) A characterization of Markov equivalence classes for acyclic digraphs, *The Annals of Statistics* 25, 505-541.

[2] Bouckaert R.R. (1995) Bayesian belief networks: from construction to evidence, PhD thesis, University of Utrecht.

[3] Chickering D.M. (2002) Optimal structure identification with greedy search, *Journal of Machine Learning Research* 3, 507-554.

[4] Neapolitan R.E. (2004) *Learning Bayesian Networks*, Pearson Prentice Hall.

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

[6] Schwarz G. (1978) Estimation the dimension of a model, *The Annals of Statistics* 6, 461-464.

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

[8] Studený M. (2005) *Probabilistic Conditional Independence Structures*, Springer-Verlag.

[9] Studený M. (2008) Mathematical aspects of learning Bayesian networks: Bayesian quality criteria, research report n. 2234, Institute of Information Theory and Automation, Prague.

[10] Studený M., Vomlel J. (2009) A reconstruction algorithm for the essential graph, *International Journal of Approximate Reasoning* 50, 385-413.

[11] Studený M., Vomlel J., Hemmecke R. (2009) A geometric view on learning Bayesian network structures, accepted in *International Journal of Approximate Reasoning*.