International Journal of Approximate Reasoning 52 (2011) 627–640

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

On open questions in the geometric approach to structural learning Bayesian nets

Milan Studený *, Jiří Vomlel

Institute of Information Theory and Automation of the ASCR, Pod Vodárenskou věží 4, Prague CZ 18208, Czech Republic

ARTICLE INFO

Article history:
Available online 27 September 2010

Keywords:
Structural learning Bayesian nets
Standard imset
Polytope
Geometric neighborhood
Differential imset

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 [18], it was shown that the set S of standard imsets is the set of vertices (=extreme points) of a certain polytope P and natural *geometric neighborhood* for standard imsets, and, consequently, for BN structures, was introduced.

The new geometric view led to a series of open mathematical questions. In this paper, 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 (=score and search method). By a *quality criterion* is meant a real function Q of a BN structure (=of a graph G, usually) and of a database D. The value $Q(G,D)$ should say how much the BN structure given by G is good to explain the occurrence of the database D. For further details about the score and search approach to structural learning Bayesian nets see [5] and recent papers [6,13].

The basic idea of an algebraic and geometric approach to this topic, proposed in Chapter 8 of [15] and then developed in [18], 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 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

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

where s_D^Q is a real number, t_D^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 Q) and $\langle *,* \rangle$ denotes the scalar product. The vector t_D^Q is named the *data vector* (relative to Q).

* Corresponding author.
E-mail addresses: studeny@utia.cas.cz (M. Studený), vomlel@utia.cas.cz (J. Vomlel).

0888-613X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2010.09.004

The main result of [18] 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 Q can be viewed as (the restriction of) an affine function on the respective Euclidean space (of higher dimension), the task to maximize 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* [11]. One of possible interpretations of this method is that it is a kind of a search method, in which one moves between 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 [18]. They are motivated by the above-mentioned intention to apply linear programming methods in the area of learning BN structures. This paper 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 $|\mathsf{X}_i| \geqslant 2$ 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 [10]. Given $G \in \mathsf{DAGS}(N)$, the respective *statistical model* of a Bayesian network structure is the class of all distributions P on X_N that factorize according to G. In this paper, we use the phrase *BN structure* (described by G) to name this statistical model.

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 standard imset into the essential graph and conversely [17].

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 $d \geqslant 1$, 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 Q on $\mathsf{DAGS}(N) \times \mathsf{DATA}(N,d)$. Given an observed database $D \in \mathsf{DATA}(N,d)$, the learning procedure based on Q consists in maximizing the function $G \mapsto Q(G,D)$ over $G \in \mathsf{DAGS}(N)$. Thus, the value $Q(G,D)$ 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 [10, Section 8.4.2].

However, there are other technical requirements on quality criteria raised in connection with computational methods for their maximization. A criterion is (additively) *decomposable* [5] if it is the sum of contributions that correspond to factors in the factorization according to the graph and *score equivalent* [3] 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) [12], but there is also a bunch of Bayesian quality criteria [16].

2.2. Elementary concepts from polyhedral geometry

Let us consider a real Euclidean space $\mathbb{R}^K$, where K is a non-empty finite set. The points in this space are vectors $\mathbf{v} = [v_s]_{s \in K}$ with $v_s \in \mathbb{R}$. The scalar product of two vectors $\mathbf{v}, \mathbf{x} \in \mathbb{R}^K$ of this type is the number

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

Given $\mathsf{V} \subseteq \mathbb{R}^K$, a convex combination of its elements is a finite sum $\sum_t \alpha_t \cdot \mathbf{v}_t$, where $\mathbf{v}_t \in \mathsf{V}$, $\alpha_t \geqslant 0$ for all t and $\sum_t \alpha_t = 1$. A set $\mathsf{V} \subseteq \mathbb{R}^K$ is called *convex* if it is closed under convex combinations and *bounded* if there exists $c > 0$ such that $-c \leqslant v_s \leqslant c$ for any $s \in K$ and $\mathbf{v} = [v_s]_{s \in K} \in \mathsf{V}$.

The set $\mathsf{P} \subseteq \mathbb{R}^K$ of all convex combinations of points in a finite set $\mathsf{V} \subseteq \mathbb{R}^K$ is called a *polytope*. If all components of vectors in V are rational numbers, that is, $\mathsf{V} \subseteq \mathbb{Q}^K$, then P is called a *rational* polytope. A special case of a polytope is a line-segment connecting vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^K$:

$$[\mathbf{x}, \mathbf{y}] \equiv \{\alpha \cdot \mathbf{x} + (1-\alpha) \cdot \mathbf{y};\ \alpha \in [0,1]\}.$$

A *polyhedron* is the set of points $\mathbf{x} \in \mathbb{R}^K$ that satisfy a finite number of linear inequality constraints, which are the requirements of the form $\langle \mathbf{v}, \mathbf{x} \rangle \leqslant \beta$, where $\mathbf{v} \in \mathbb{R}^K$ and $\beta \in \mathbb{R}$. A linear equality constraint is then the requirement $\langle \mathbf{v}, \mathbf{x} \rangle = \beta$.[1]

[1] Actually, this is a pair of inequality constraints: $\langle \mathbf{v}, \mathbf{x} \rangle \leqslant \beta$ and $\langle -\mathbf{v}, \mathbf{x} \rangle \leqslant -\beta$.

A well-known result in polyhedral geometry says that a set $\mathsf{P} \subseteq \mathbb{R}^K$ is a polytope iff it is a bounded polyhedron [11, Corollary 7.1c]. Note that the classic version of the *simplex method* is applicable to the task to find maximum/minimum of a linear function $\mathbf{x} \mapsto \langle \mathbf{v}, \mathbf{x} \rangle$ over $\mathbf{x} \in \mathsf{P}$, where P is a polyhedron and $\mathbf{v} \in \mathbb{R}^K$ [11, Chapter 11].

A *vertex* (=an extreme point) of a polytope P is a point $\mathbf{x} \in \mathsf{P}$ which cannot be written as a convex combination of elements in $\mathsf{P} \setminus \{\mathbf{v}\}$.[2] An *edge* of a polytope P is a line-segment $[\mathbf{x}, \mathbf{y}]$, where $\mathbf{x}$, $\mathbf{y}$ are distinct vertices of P and the set $\mathsf{P} \setminus [\mathbf{x}, \mathbf{y}]$ 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 edges, for details see [11, Section 11.1].

A *conical combination* of elements in $\mathsf{V} \subseteq \mathbb{R}^K$ is a finite sum $\sum_t \alpha_t \cdot \mathbf{v}_t$, where $\mathbf{v}_t \in \mathsf{V}$ and $\alpha_t \geqslant 0$ for all t. A set $\mathsf{C} \subseteq \mathbb{R}^K$ is a cone if it is closed under conical combinations. A *rational polyhedral cone* is the set C of all conical combinations of points in a finite set $\mathsf{V} \subseteq \mathbb{Q}^K$; we say then that C is spanned by V. A cone C is *pointed* if there exists non-zero $\mathbf{v} \in \mathbb{R}^K$ such that $\langle \mathbf{v}, \mathbf{x} \rangle > 0$ for any non-zero $\mathbf{x} \in \mathsf{C}$.

2.3. Imsets

The method of *structural imsets* was proposed in [15] 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 $\mathbb{Z}^{\mathcal{P}(N)}$, called a *lattice point* in $\mathbb{R}^{\mathcal{P}(N)}$ in the context of integer programming [11].

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

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

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 with integers as coefficients.

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

An imset will be called *combinatorial* if it is a linear combination of elementary imsets with non-negative integers as coefficients.[3] 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 [15, Proposition 4.3] 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.

An imset which is a combination of elementary imsets with non-negative rational coefficients will be called *structural*.[4] Note that these imsets are the tools for describing probabilistic conditional independence structures [15]. There exists a structural imset over N with $|N| = 5$ which is not combinatorial [8].

2.4. Algebraic approach to learning BN structures

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

$$u_G = \delta_N - \delta_\emptyset + \sum_{i \in N} \left\{ \delta_{\mathrm{pa}_G(i)} - \delta_{\{i\} \cup \mathrm{pa}_G(i)} \right\}, \tag{2}$$

where $\mathrm{pa}_G(i) = \{j \in N; j \to i \text{ in } G\}$ denotes the set of *parents* of i in G. Note that 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. Hence, the memory demands for representing standard imsets in a computer are polynomial in $|N|$.

[2] Equivalently, $\mathsf{P} \setminus \{\mathbf{v}\}$ is convex.

[3] Equivalently, a combinatorial imset is a sum of elementary imsets with allowed repetition of summands.

[4] An equivalent characterization is that a structural imset is a lattice point within the cone $\mathcal{R}(N)$.

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 [15, Corollary 7.1]. In particular, the standard imset for $G \in \mathsf{DAGS}(N)$ is a unique (algebraic) 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 Section 5 of this paper).[5] The degree of a standard imset u_G for $G \in \mathsf{DAGS}(N)$ equals $\binom{|N|}{2} - r$, where r is the number of arrows in G [15, Lemma 7.1].

Now, Lemmas 8.3 and 8.7 from [15] together say that every score equivalent and decomposable criterion Q must have the form

$$Q(G,D) = s^Q_D - \langle t^Q_D, u_G \rangle \quad \text{for } G \in \mathsf{DAGS}(N), D \in \mathsf{DATA}(N,d), \quad d \geqslant 1, \tag{3}$$

where the constant $s^Q_D \in \mathbb{R}$ and the vector $t^Q_D \in \mathbb{R}^{\mathcal{P}(N)}$ do not depend on G. The formulas for the data vector t^Q_D relative to some basic quality criteria Q have been derived in [15,16].

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:

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

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 describing the same BN structure. Theorem 4 in [18] says that S is the set of vertices of a rational polytope $\mathsf{P} \subseteq \mathbb{R}^{\mathcal{P}(N)}$. This polytope P will be called the *standard imset polytope* in the sequel. It follows from (3) that the task to maximize Q over $G \in \mathsf{DAGS}(N)$ is equivalent to the task to minimize the linear function $u \mapsto \langle t^Q_D, u \rangle$ 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 *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 [18, Theorem 5] that the well-known *inclusion neighborhood*, used widely in current computational methods for learning BN structures, like the GES algorithm [5], is strictly contained in the geometric one. Moreover, it follows from [15, Corollary 8.4] that standard imsets $u, v \in \mathsf{S}$ correspond to inclusion neighbors iff their difference $w = u - v$ is either an elementary imset or its multiple by -1.

The importance of the concept of geometric neighborhood is based on the fact that, for any affine function Q on $\mathbb{R}^{\mathcal{P}(N)}$, a local maximum of Q in $u \in \mathsf{S}$ with respect to the geometric neighborhood must be the global maximum of Q over P [18, Theorem 6]. In particular, this holds for any reasonable quality criterion Q for learning BN structures.

The following research goals have been expressed in Conclusions of [18].

- Describe the linear constraints on the points in P. A complete characterization of these constraints would provide a polyhedral description of P, required by the classic version of the simplex method.
- An interesting conjecture by Raymond Hemmecke was that the only lattice points within P are standard imsets.
- Describe the *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 maximizing a quality criterion Q. In this paper we partially answer 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 the points in P.

3.1. Overview of the constraints

We classify our linear constraints into three groups, denoted (A), (B) and (C). First, the fact that every standard imset is combinatorial [15, Lemma 7.1] implies that it belongs to the cone $\mathcal{R}(N)$ generated by elementary imsets. This simple observation gives two kinds of necessary linear conditions on the points in 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[6]:

[5] Conversely, every elementary imset is the standard one for some $G \in \mathsf{DAGS}(N)$.

[6] [15, Proposition 4.4] says that every structural imset is o-standardized, which basically means that (A.1) and (A.2) are valid, cf. [15, p. 40].

Table 1
Numbers of non-specific inequality constraints.

$\lvert N\rvert$	2	3	4	5
Number of extreme rays	1	5	37	117978

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

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

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

Specifically, we will interpret (A.1) and (A.2) in this way: $u\in \mathsf{S}$ is uniquely determined by its values $u(S)$ for $S\subseteq N$ with $|S|\geqslant 2$. The remaining values are then determined in two steps as follows:

- $\forall a\in N \qquad u(\{a\})=-\sum_{S,a\in S,|S|\geqslant 2}u(S)$,
- $u(\emptyset)=-\sum_{S,S\neq\emptyset}u(S)$.

In the sequel, we are interested in describing standard imsets in this way and formulate the other constraints accordingly.

(B) Non-specific inequality constraints

The inequality constraints on the 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

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

An equivalent definition is that $\langle m,v\rangle\geqslant 0$ for every elementary imset v over N, see [15, Proposition 5.1]. This observation gives a (formally infinite) set of inequality constraints on the points in $\mathcal{R}(N)$, and, therefore, on any standard imset u:

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

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 $|S|\leqslant 1$. Second, the class of these special supermodular functions is a pointed rational polyhedral cone and has, therefore, finitely many extreme rays.[7] Every such extreme ray has unique lattice point representative with the property that its components have no common prime divisor. Thus, the class of these representatives, denoted by $\mathcal{K}_\ell^\diamond(N)$ and called the ℓ-skeleton in [15], establishes a finite set of normalized inequality constraints:

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

These (representatives of) extreme rays have been computed for $|N|\leqslant 5$ using linear programming packages [14]. It seems that the number of these extreme rays grows super-exponentially with $|N|$; their numbers for $|N|\leqslant 5$ are in Table 1. It looks like none of these inequality constraints on points in P is derivable from the other constraints (including those mentioned in subsequent subsection (C)).

A standard example of a supermodular function is the identifier $m^{A\uparrow}$ of the class of supersets of $A\subseteq N$ given by:

$$m^{A\uparrow}(S)=\begin{cases}1 & \text{if } A\subseteq S,\\ 0 & \text{otherwise.}\end{cases}$$

Thus, an easy consequence of the condition (B) is the following constraint:

$$\forall A\subseteq N \quad \langle m^{A\uparrow},u\rangle\equiv\sum_{S,A\subseteq S}u(S)\geqslant 0. \tag{4}$$

However, if the constraints (A) are valid, (4) is non-trivial only for sets A with $|A|\geqslant 2$.

(C) Specific inequality constraints

The results of [17] led to a series of specific linear inequality constraints on standard imsets, that is, the constraints 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

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

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 constraints:

[7] See [15, Section 5.1.2 and Lemma 5.3] for both these claims.

Table 2
Numbers of specific inequality constraints.

$\|N\|$	2	3	4	5
Number of classes	4	18	166	7579

$$\sum_{S\in\mathcal{A}} u(S) \leqslant 1 \quad \text{for any system } \emptyset \neq \mathcal{A} \subseteq \{S \subseteq N;\ |S| \geqslant 1\} \text{ closed under supersets.} \tag{C}$$

Note that, unlike the number (B)-constraints, the number of constraints in (C) seems to grow only exponentially with $|N|$; their numbers for $|N| \leqslant 5$ are given in Table 2. Actually, these constraints are in correspondence with hierarchical log-linear models over N. 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* [9] also correspond to classes of incomparable subsets of the class $\{A \subseteq N; |S| \geqslant 1\}$, namely to those whose union is N.

Nevertheless, the list of conditions in (C) is not reduced completely. Some of these constraints are superfluous because they follow from the other ones combined with (A) and (B).[8] More specifically, after the reduction one has 1 specific inequality constraint for $|N| = 2$, 8 constraints for $|N| = 3$ and 117 for $|N| = 4$. Moreover, each of the (C)-constraints can be, owing to (A.2), reformulated equivalently in a "normalized" form

$$\sum_{S\in\mathcal{B}} k_S \cdot u(S) \leqslant 1 \quad \text{for some } \mathcal{B} \subseteq \{S \subseteq N;\ |S| \geqslant 2\} \quad \text{and} \quad k_S \in \mathbb{Z} \setminus \{0\}.$$

It looks like none of the constraints for $\mathcal{A} \subseteq \{S \subseteq N;\ |S| \geqslant 2\}$ is superfluous, while if $\mathcal{A}$ contains a singleton then both cases can occur: either the respective inequality constraint is superfluous or it is non-derivable from the others.[9]

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

Proof. The proof is based on some results from [17]. One can proceed by the induction on $|N|$. If N is a singleton then, by (A.1) and (A.2), the only $u \in \mathsf{S}$ is the zero imset, which satisfies (C).

If $|N| \geqslant 2$, we use the reduction steps from [17, Section 6]. Let us exclude the trivial case of the zero imset. A non-zero imset $u \in \mathsf{S}$ is either adapted, which means $u(N) \neq 0$, or it is not adapted, that is, $u(N) = 0$.

- If u is not adapted then Corollary 5.1 and Lemma 6.2 in [17] imply there exists a set $\emptyset \neq M_u \subset N$, called the core of u, such that $u(S) = 0$ for $S \subseteq N$ with $S \setminus M_u \neq \emptyset$ and the restriction of u to $\mathcal{P}(M_u)$ is a standard imset over M_u. Now, given a system $\mathcal{A}$ from (C), $\mathcal{A} \cap \mathcal{P}(M_u)$ is a system of subsets of M_u closed under supersets, and, by the induction hypothesis,

$$\sum_{S\in\mathcal{A}} u(S) = \sum_{S\in\mathcal{A}\cap\mathcal{P}(M_u)} u(S) \leqslant 1,$$

which verifies the induction step.

- If u is adapted then [17, Lemma 6.3 including preceding explanation] implies that there exist sets $\emptyset \neq M, T \subset N$ such that $M \cup T = N$, the imset $\tilde{w} \equiv u - \delta_N + \delta_T + \delta_M - \delta_{T\cap M}$ vanishes for $S \subseteq N$ with $S \setminus M \neq \emptyset$ and the restriction w of $\tilde{w}$ to $\mathcal{P}(M)$ is a standard imset over M. That means

$$u = \tilde{w} + \underbrace{\delta_N - \delta_T - \delta_M + \delta_{T\cap M}}_{v}.$$

Given a class $\mathcal{A}$ from (C), the set system $\mathcal{A} \cap \mathcal{P}(M)$ is a class of subsets of M closed under supersets, and, by the induction hypothesis one has

$$\sum_{S\in\mathcal{A}} \tilde{w}(S) = \sum_{S\in\mathcal{A}\cap\mathcal{P}(M)} w(S) \leqslant 1.$$

To verify the induction step we distinguish two subcases:

– If $M \notin \mathcal{A}$, then, since $\mathcal{A}$ is closed under supersets, $\mathcal{A} \cap \mathcal{P}(M)$ is empty, which even says $\sum_{S\in\mathcal{A}} \tilde{w}(S) = \sum_{S\in\mathcal{A}\cap\mathcal{P}(M)} w(S) = 0$. As $T \cap M \notin \mathcal{A}$, the special form of the imset $v = u - \tilde{w}$ implies that $\sum_{S\in\mathcal{A}} v(S) = \sum_{S\in\mathcal{A}} \delta_N(S) - \delta_T(S)$ is either 1 or 0 (as $N \in \mathcal{A}$). Thus, $\sum_{S\in\mathcal{A}} v(S) \leqslant 1$ and, by summing, $\sum_{S\in\mathcal{A}} u(S) = \sum_{S\in\mathcal{A}} \tilde{w}(S) + \sum_{S\in\mathcal{A}} v(S) \leqslant 0 + 1$ one gets what is desired.

[8] For example, if $\mathcal{A} = \{S \subseteq N;\ a \in S\}$ for some $a \in N$, then (A.2) gives $\sum_{S\in\mathcal{A}} u(S) = 0 \leqslant 1$.
[9] The latter case may happen for $|N| = 5$.

– If $M \in \mathcal{A}$ then it is enough to show $\sum_{S\in\mathcal{A}} v(S) \leqslant 0$. One can either have $T \cap M \in \mathcal{A}$, in which case $T, M, N \in \mathcal{A}$ and $\sum_{S\in\mathcal{A}} v(S) = 1 - 1 - 1 + 1 = 0$, or one can have $T \cap M \notin \mathcal{A}$, in which case one writes

$$\sum_{S\in\mathcal{A}} v(S) = \sum_{S\in\mathcal{A}} \underbrace{\delta_N(S)}_{1} - \delta_T(S) - \underbrace{\delta_M(S)}_{1} = \sum_{S\in\mathcal{A}} -\delta_T(S) \leqslant 0.$$

This concludes the proof. □

3.2. Conjecture about the linear constraints

The constraints (A)–(C) from Subsection 3.1 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: $u(S) \geqslant -1$ for any $S \subseteq N$. These consequences are formulated in Appendix A.

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

Conjecture 1. *The linear constraints (A)–(C) together give a necessary and sufficient condition on a vector $u \in \mathbb{R}^{\mathcal{P}(N)}$ to belong to P.*

It seems to be a challenge to existing software either to confirm or disprove Conjecture 1 for $|N| = 5$. Actually, for this reason, we do not know the exact number of (C)-constraints (=after a complete reduction) in 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 $|N| \leqslant 5$ and the result was negative. This led him to a hypothesis that every lattice point in the standard imset polytope is already a standard imset. In this paper, we confirm the hypothesis as a consequence of the inequality constraints from Section 3.

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

Proof. Let us denote $\mathcal{P}_*(N) = \{A \subseteq N; \mid A \mid \geqslant 2\}$. The basic idea is to introduce a special linear transformation L from $\mathbb{R}^{\mathcal{P}_*(N)}$ to $\mathbb{R}^{\mathcal{P}_*(N)}$. More specifically, any $u \in \mathbb{R}^{\mathcal{P}_*(N)}$ is assigned $L(u) = v \in \mathbb{R}^{\mathcal{P}_*(N)}$ by the formula

$$v(S) = \sum_{T, S\subseteq T\subseteq N} u(T) \quad \text{for } S \in \mathbb{R}\mathcal{P}_*(N). \tag{5}$$

To show that L is a one-to-one mapping define a mapping M which maps $v \in \mathbb{R}^{\mathcal{P}_*(N)}$ to $M(v) = w \in \mathbb{R}^{\mathcal{P}_*(N)}$:

$$w(A) = \sum_{S, A\subseteq S\subseteq N} (-1)^{|S\setminus A|} \cdot v(S) \quad \text{for } A \in \mathbb{R}\mathcal{P}_*(N). \tag{6}$$

To see that M is the inverse of L, given $A \in \mathbb{R}\mathcal{P}_*(N)$, we substitute (5) into (6) and change the order of summation:

$$w(A) = \sum_{S, A\subseteq S\subseteq N} (-1)^{|S\setminus A|} \cdot \sum_{T, S\subseteq T\subseteq N} u(T) = \sum_{T, A\subseteq T\subseteq N} u(T) \cdot \sum_{S, A\subseteq S\subseteq T} (-1)^{|S\setminus A|} = \sum_{T, A\subseteq T\subseteq N} u(T) \cdot \delta_A(T) = u(A).$$

Thus, L is a one-to-one linear mapping. Moreover, it has the property that both L and its inverse M maps lattice points to lattice points: this follows from (5) and (6).

Nevertheless, L can be viewed as a linear mapping from the linear subspace of $\mathbb{R}^{\mathcal{P}(N)}$ specified by (A.1) and (A.2) (see Section 3.1). This linear subspace includes the polytope P. Since a linear mapping preserves convex combinations, the image of P by L is a polytope whose vertices are images of vertices of P, that is, of standard imsets. Now, as explained in subsection (B) of Section 3.1, the condition (4) is valid for any standard imset. In particular, given $u \in \mathsf{S}$, its image $L(u)$ has non-negative components $\langle m^{A\uparrow}, u\rangle$ for $A \subseteq N$, $|A| \geqslant 2$. Moreover, as $\mathcal{A} = \{T; A \subseteq T \subseteq N\}$ is closed under supersets, one has, by (C), $L(u)(A) = \sum_{S\in\mathcal{A}} u(S) \leqslant 1$. In particular, $L(u)(A) \in \{0,1\}$ for any $A \subseteq N$, $|A| \geqslant 2$, which means L maps S into $\{0,1\}^{\mathcal{P}_*(N)}$. There is no lattice point in the interior of the hypercube $[0,1]^{\mathcal{P}_*(N)}$, which implies the same holds for P. □

Remark. The reader may be interested in the history of proving Theorem 1. The original proof, which we had mentioned in our WUPES 2009 contribution, was quite long and complicated. It was based on technical details of the reconstruction algorithm from [17]. However, later e-mail discussion with Raymond Hemmecke and Silvia Lindner, in connection with another paper [4], had an important side-effect. All of us have realized the potential of the transformation L given by (5), which makes it possible to simplify things substantially. Actually, our further (joint) research topic will be whether the transformation L (possibly followed by another one) can lead to a simpler algebraic representative of a BN structure [19]. Another comment is that the key tools in our final proof, the transformation given by (5) and its inverse (6), well-known in much more general context as *Möbius inversion* [2], has numerous applications in the area of uncertainty processing in artificial intelligence, see the survey [7].

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

Conjecture 2. *The constraints (A)–(C) together form a necessary and sufficient condition on an imset $u \in \mathbb{Z}^{\mathcal{P}(N)}$ to be a standard imset (over N).*

Indeed, if Conjecture 1 is true then, by Theorem 1, Conjecture 2 holds as well. However, it is not clear at this moment whether the proof of Conjecture 2 is enough to confirm Conjecture 1. Perhaps the polyhedron specified by inequalities (A)–(C) has, for some $|N| \geqslant 5$, vertices that are not lattice points.

On the other hand, Conjecture 2 could be easier to prove. Actually, we believe we can offer a method to verify a hypothesis like that. The idea is to use an algorithmic characterization of standard imsets from [17, Section 7.4]. This seems to allow one to verify the conjecture by induction on $|N|$; for details see Section A. However, it may be the case that (A)–(C) is not a complete list of linear constraints on S for $|N| \geqslant 5$. Then additional constraints may be added to that list and one can try to verify the modified conjecture by the same method.

5. A catalogue of differential imsets over 4 variables

The result of our analysis of the geometric neighborhood in the case $|N| = 4$ is an electronic catalogue. Actually, it is a catalogue of differential imsets for pairs of geometric neighbors. 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 $u, v \in \mathsf{S}$ yielding w. This number seems to be quite important characteristic of w.

We say that two imsets u, v over N are *permutation equivalent* (PE) if there exists a bijection $\pi : N \to N$ such that, for all $A \subseteq N$, it holds $u(A) = v(\pi(A))$, where $\pi(A) = \{\pi(i); i \in A\}$. 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 differential imsets. An elegant solution is offered below.

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

Proof. Consider an auxiliary notation: given a triplet of sets $\langle i,B|C\rangle$, where B, $C \subseteq N$ are disjoint and $i \in N \setminus B \cup C$, we introduce[10]:

$$u_{\langle i,B|C\rangle} = \delta_{\{i\}\cup B\cup C} + \delta_C - \delta_{\{i\}\cup C} - \delta_{B\cup C}.$$

Observe that $u_{\langle i,B|C\rangle}$ is a sum of (distinct) elementary imsets of the form $u_{\langle i,j|K\rangle}$, where $j \in B$ and $C \subseteq K \subseteq N$. Indeed, if $B = \emptyset$ then $u_{\langle i,B|C\rangle} = 0$ and if $B \neq \emptyset$ then we order its elements in a sequence $j_1, \ldots, j_\ell$, $\ell \geqslant 1$ and get

$$u_{\langle i,B|C\rangle} = u_{\langle i,j_1|C\rangle} + u_{\langle i,j_2|C\cup\{j_1\}\rangle} + \cdots + u_{\langle i,j_\ell|C\cup\{j_1,\ldots,j_{\ell-1}\}\rangle}.$$

Given $G \in \mathsf{DAGS}(N)$, let us fix an ordering σ of its nodes that is consonant with the direction of arrows in G.[11] It has been shown in [15, proof of Lemma 7.1] that the standard imset u_G can be written as follows:

$$u_G = \sum_{i\in N} u_{\langle i,\mathrm{pre}_{\sigma,G}(i)|\mathrm{pa}_G(i)\rangle},$$

where $\mathrm{pre}_{\sigma,G}(i)$ denotes the set of predecessors of i in σ that are not in $\mathrm{pa}_G(i)$. As mentioned above, each summand $u_{\langle i,\mathrm{pre}_{\sigma,G}(i)|\mathrm{pa}_G(i)\rangle}$ is a sum of distinct elementary imsets $u_{\langle i,j|K\rangle}$ where j precedes i in σ. These groups of elementary imsets are disjoint for distinct i's. Thus, u_G can be written as a sum of distinct elementary imsets (=a combination with coefficients + 1 and 0). □

[10] In the triplets, we use a shorthand i to denote $\{i\}$.

[11] That means, if $a \to b$ in G then a precedes b in σ.

A 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.*

A differential imset $w = u - v$ for $u,v \in \mathsf{S}$ is a combinatorial imset iff it is a (possibly empty) combination of elementary imsets with coefficients +1, that is, the sum of elementary imsets without repetition; the number of summands is deg(u) − deg(v) then.

Proof. The first claim is evident from Lemma 2. To prove the remaining claims the following graphical view is useful. Let us imagine the "inclusion neighborhood" graph: an undirected graph over S in which $u, v \in \mathsf{S}$ are adjacent nodes iff they correspond to inclusion neighbors. Then $u - v$ is (±1)-multiple of an elementary imset (see Section 2.5) and the degree difference $\deg(u) - \deg(v)$ is either +1 or −1 (cf. Section 2.3). In particular, the graph is special: its nodes are grouped into layers that correspond to the degree of standard imsets and edges are only between successive layers.

An important auxiliary observation is this: given $K, L \in \mathsf{DAGS}(N)$ such that $u_L - u_K$ is a non-zero combinatorial imset there exists an "ascending" path $u_K = v_1, v_2, \ldots, v_\ell = u_L$, $\ell \equiv \deg(u_L) - \deg(u_K) + 1$ in the above graph, that is, v_{i+1} is one layer above v_i for $i = 1, \ldots, \ell - 1$. Indeed, the premise means that the statistical model of a BN structure (see Section 2.1) described by L is contained in the one described by K [15, Lemma 8.6]. Then Chickering's transformational characterization of this inclusion [5] says there exists a sequence of $K = G_1, \ldots, G_n = L \in \mathsf{DAGS}(N)$, $n \geqslant 2$ such that $\forall\ i = 1, \ldots, n-1$ either G_{i+1} gives the same statistical model as G_i or G_{i+1} is obtained from G_i by a removal of one arrow (cf. [15, Lemma 8.5]). If G_{i+1} and G_i give the same statistical model then $u_{G_{i+1}} = u_{G_i}$ (see Section 2.4). If G_{i+1} is obtained by the removal of an arrow $a \to b$ in G_i, then, by [15, Proposition 8.3], $u_{G_{i+1}} - u_{G_i}$ is an elementary imset of the form $u_{\langle a,b|C\rangle}$ where $C \subseteq N \setminus \{a,b\}$. That means, $u_{G_{i+1}}$ is an inclusion neighbor of u_{G_i} which is one layer above it. Thus, the existence of the ascending path was verified. Moreover, the summands in the corresponding decomposition $u_L - u_K = \sum_{i=1}^{\ell-1}(v_{i+1} - v_i)$ must be distinct elementary imsets, because they (graphically) correspond to the removals of different arrows.

This already implies the third claim. If $w = u - v$ is a non-zero combinatorial imset then the above-mentioned ascending path from v to u of the length $\deg(u) - \deg(v)$ defines the desired decomposition of w into distinct elementary imsets. If w is the zero imset, it is an empty combination of elementary imsets with coefficients +1. The converse implication in the third claim is trivial.

To show the second claim, about the maximal number $\binom{|N|}{2}$ of summands, assume without loss of generality that neither $w = u - v$ nor $-w = v - u$ is a combinatorial imset, for otherwise the above claim already implies what is required. Given any $u \in \mathsf{S}$, the auxiliary observation above implies there exists an ascending path from the zero imset 0 to u and an ascending path from u to the the imset u^* of the highest degree.[12] If we denote $m \equiv \deg(u^*) = \binom{|N|}{2}$ then the former path has the length $\deg(u)$ and the latter $m - \deg(u)$; their concatenation has the length m.

Thus, given distinct $u,v \in \mathsf{S}$, consider both an ascending path ρ_u from 0 to u^* through u and an analogous path ρ_v through v. Now, the upper part of ρ_v (=the section from v to u^*) cannot share any node with lower part of ρ_u (=the section from 0 to u) for otherwise an ascending path from v to u exists in the inclusion neighborhood graph, which contradicts the assumption that $w = u - v$ is not a combinatorial imset. Analogously, the upper part of ρ_u does not meet the lower part of ρ_v. In particular, the concatenation of ρ_u and ρ_v is a pseudo-cycle of the length $2m$, which can be shortened to a cycle containing both u and v (of the length at most $2m$). Thus, at least one part of the cycle is a path between u and v of the length at most m. Moreover, the obtained path between u and v consists of two monotone parts (in sense of the degree).

Therefore, the above path $u = u_1, u_2, \ldots, u_t = v$, $2 \leqslant t \leqslant m + 1$ defines a decomposition of the corresponding differential imset

$$w \equiv u - v = \sum_{i=1}^{t-1} \{u_i - u_{i+1}\},$$

where the summands are either elementary imsets or their multiples by −1. This proves the second claim. □

In particular, every differential imset for a pair of geometric neighbors can be expressed in the described way, which we have utilized 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 the imsets with a non-negative degree difference are kept there.

[12] This is the standard imset corresponding to the empty graph.

Table 3
Numbers of pairs of geometric neighbor, differential imsets and PE classes.

Degree difference	Pairs of neighbors	Differential imsets	PE classes
0	2894	927	88
1	4248	1359	144
2	1296	505	71
3	80	40	16
Total	8518	2831	319

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 degree differences (for geometric neighbors) are integers between 0 and 3. The values of the squared Euclidean length are even numbers between 4 and 22. The numbers of non-zero imset values are integers between 4 and 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 3 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

5.3. A simple example

As mentioned in Section 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 3, this is not true for $|N| = 4$: there are 144 PE classes with the degree difference 1 while one has only 3 PE classes consisting of elementary imsets.

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

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

where

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

5.4. Some preliminary observations

We used the catalogue of differential imsets referenced above to perform a preliminary analysis of the geometrical neighborhood for $|N| = 4$. The first observation, somewhat surprising, was how different can be two geometrical neighbors graphically. In the catalogue there are eight PE classes of differential imsets that corresponds to geometrical neighbors that differ by as much as six edges. Their respective numbers in the catalogue are 29, 33, 35, 69, 255, 270, 286, and 292. At the same time, these eight differential imsets are combinations of six elementary imsets with coefficients +1 and −1, which shows that the upper bound $\binom{|N|}{2}$ in Lemma 3 is tight for $|N| = 4$.

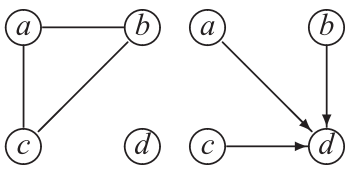

Fig. 1. An example of geometrical neighbors whose essential graphs differ by six edges.

In Fig. 1 we give an example of two essential graphs of (standard imsets that are) geometric neighbors and differ by six edges. Their differential imset is

$$\delta_{\{a\}} + \delta_{\{b\}} - 2\delta_{\{d\}} - \delta_{\{a,b\}} + \delta_{\{c,d\}} - \delta_{\{a,b,c\}} + \delta_{\{a,b,d\}} = -u_{\langle a,b|\emptyset\rangle} + u_{\langle a,d|\emptyset\rangle} + u_{\langle c,d|ab\rangle} - u_{\langle b,c|ad\rangle} + u_{\langle b,d|a\rangle} - u_{\langle a,c|d\rangle}.$$

For $|N| = 4$, the maximal number of pairs of standard imsets that are geometrical neighbors and give rise to the same differential imset is sixteen. The corresponding differential imset[13] is the elementary imset $u_{\langle b,c|a\rangle}$, which has the number 89 in the catalogue. This means geometrically that there are sixteen parallel edges of the standard imset polytope.

6. Conclusions

Let us mention some of our research goals motivated by the results reported here. First, we would like either to confirm or disprove Conjecture 1 from Section 3.2 for $|N| = 5$. If this is confirmed for $|N| = 5$ we may try to verify its weaker version Conjecture 2 from Section 4 for general $|N|$, by the method described in Appendix A.

The catalogue from Section 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 the geometric neighborhood, namely whether differential imsets (for geometric neighbors) correspond to some graphical operations with the corresponding essential graphs. However, the example from Fig. 1 suggests that this may be quite difficult task. Indeed, these two graphs differ completely in their adjacencies, and, despite that, they represent geometric neighbors.

Lemma 3, namely its second claim, may appear to be an important step to show that the inclusion of Bayesian nets can be tested with polynomial complexity in $|N|$ by an algebraic method proposed in [4].

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 MŠMT n. 2C06019 and Eurocores LogICCC n. ICC/08/E010. We thank the reviewers for their valuable comments.

Appendix A. Basic consequences of the constraints

In this section, we gather basic consequences of the constraints (A)–(C) from Section 3. Then we explain what remains to be done to verify Conjecture 2 by a method based on the procedure for testing standard imsets from [17].

Definition 1 (*some basic terminology*).
Given a non-zero $u \in \mathbb{Z}^{\mathcal{P}(N)}$ satisfying (A)–(C), the set

$$M_u \equiv \bigcup \{S \subseteq N;\ u(S) \neq 0\},$$

will be called the *core* of u. If $M_u = N$ then u will be called *adapted*. The class

$$\mathcal{T}_u = \{T \subseteq N;\ u(T) < 0\},$$

will be called the class of *negative sets* (in u) and the class

$$\mathcal{L}_u = \{L \subseteq N;\ u(L) > 0\ \&\ \emptyset \neq L \neq M_u\},$$

the class of *positive sets* (in u).[14] The symbols $\mathcal{T}_u^{max}$ and $\mathcal{L}_u^{max}$ will denote the classes of maximal sets (with respect to inclusion) in $\mathcal{T}_u$ and $\mathcal{L}_u$, respectively.

Lemma 4 (*basic observations*). *Let $u \in \mathbb{Z}^{\mathcal{P}(N)}$ satisfy (A)–(C). If $u \neq 0$ then $u(M_u) = 1$ for the core M_u. In particular, $u(N) \in \{0,1\}$. Moreover, $u(\emptyset) \geqslant 0$ and*

$$\forall S \subseteq T \subseteq N \quad \sum_{K, S \subseteq K \subseteq T} u(K) \geqslant -1. \tag{†}$$

In particular, $u(S) \geqslant -1$ for every $S \subseteq N$.

Proof. Let us put

$$\mathcal{M}_u = \{L \subseteq N; \exists T \subseteq N\ L \subseteq T\ \&\ u(T) \neq 0\}.$$

Thus, $u \neq 0 \Rightarrow M_u = \bigcup \mathcal{M}_u \neq \emptyset$. Indeed, if $u(\emptyset) \neq 0$ then, by (A.1), there exists non-empty $T \subseteq N$ such that $u(T) \neq 0$ and $\mathcal{M}_u$ contains at least one non-empty set.

[13] It is unique up to its permutations.

[14] Let us emphasize that, by definition, we exclude both the core M_u and the empty set $\emptyset$ from the class of positive sets $\mathcal{L}_u$.

First, we observe $S \in \mathcal{M}_u^{max} \Rightarrow u(S) \geqslant 1$. Indeed, the constraint (B) implies (4), which gives $u(S) = \sum_{T, S \subseteq T} u(T) \geqslant 0$. But $0 \neq u(S) \in \mathbb{Z}$ implies $u(S) \geqslant 1$ then.

Second, observe that $\mathcal{M}_u$ has a unique maximal set. Put $\mathcal{A} = \{T \subseteq N;\ \exists S \in \mathcal{M}_u^{max}\ S \subseteq T\}$ and write by (C), respectively by (A.1):

$$|\ \mathcal{M}_u^{max}\ | \leqslant \sum_{S \in \mathcal{M}_u^{max}} u(S) = \sum_{S \in \mathcal{A}} u(S) \leqslant 1.$$

Thus, the unique maximal set in $\mathcal{M}_u$ is the core M_u. We already know $u(M_u) \geqslant 1$ and the converse inequality follows from (C), respectively (A.1): $u(M_u) = \sum_{S \in \mathcal{A}} u(S) \leqslant 1$.

The fact $u(\emptyset) \geqslant 0$ follows from (B) and from the observation that $\delta_\emptyset$ is a supermodular function: $u(\emptyset) = \langle \delta_\emptyset, u \rangle \geqslant 0$.

To see (†) realize that $\sum_{K, S \subseteq K} u(K) \geqslant 0$ by (4). Observe that the class

$$\mathcal{A} = \{K;\ S \subseteq K\ \&\ \neg(K \subseteq T)\} = \{K;\ S \subseteq K\ \&\ K \cap (N \setminus T) \neq \emptyset\}$$

is closed under supersets. Therefore, by (C), $\sum_{K \in \mathcal{A}} u(K) \leqslant 1$, which can be re-written as $-\sum_{K \in \mathcal{A}} u(K) \geqslant -1$. Hence,

$$\sum_{K, S \subseteq K \subseteq T} u(K) = \sum_{K, S \subseteq K} u(K) - \sum_{K \in \mathcal{A}} u(K) \geqslant 0 - 1 \geqslant -1.$$

If one has $T = S$ then (†) reduces to $u(S) \geqslant -1$. □

Corollary 1. *If $0 \neq u \in \mathbb{Z}^{\mathcal{P}(N)}$ satisfies (A)–(C) then any $T \in \mathcal{T}_u$ is non-empty and $\mathcal{T}_u = \{T \subseteq N;\ u(T) = -1\}$.*

Proof. By Lemma 4, $u(\emptyset) \geqslant 0$ gives $\emptyset \notin \mathcal{T}_u$ and $u(T) \geqslant -1$ for $T \subseteq N$ implies the rest. □

Definition 2 (center). Given $0 \neq u \in \mathbb{Z}^{\mathcal{P}(N)}$ satisfying (A)–(C) let us introduce

$$W_u \equiv \{a \in N; \exists\ T, T' \in \mathcal{T}_u,\ T \neq T',\ a \in T \cap T'\}$$

and call it the *center* of u.

Lemma 5 (about the center).
Assume that $0 \neq u \in \mathbb{Z}^{\mathcal{P}(N)}$ satisfies (A)–(C). Then $M_u = \bigcup \mathcal{T}_u = \bigcup \mathcal{T}_u^{max}$ and $W_u = \bigcup \mathcal{L}_u = \bigcup \mathcal{L}_u^{max}$.

Proof. The inclusion $\bigcup \mathcal{T}_u \subseteq M_u$ follows from the definition of the core M_u. To see the converse inclusion use the condition (A.2):

$$\forall a \in M_u \qquad 0 = \sum_{S, a \in S} u(S) = \underbrace{\sum\nolimits_{S, a \in S \& u(S) > 0} u(S)}_{\geqslant 1} + \sum_{S \in \mathcal{T}_u, a \in S} u(S),$$

where we used $u(M_u) = 1$ (Lemma 4). Hence, there exists $T \in \mathcal{T}_u$ with $a \in T$.

To show $W_u \subseteq \bigcup \mathcal{L}_u$ consider $a \in W_u$. This means

$$\sum_{R, a \in R\ \&\ u(R) < 0} u(R) \leqslant -2 \quad \text{and, by (A.2),} \qquad \sum_{L, a \in L\ \&\ u(L) > 0} u(L) \geqslant 2.$$

As $u(M_u) = 1$, there exists $L \neq M_u$ with $u(L) > 0$ and $a \in L$. Thus, $L \in \mathcal{L}_u$ and $a \in L \subseteq \bigcup \mathcal{L}_u$.

To show the converse inclusion $\bigcup \mathcal{L}_u \subseteq W_u$ consider $a \in \bigcup \mathcal{L}_u$. Thus, $L \in \mathcal{L}_u$ with $a \in L$ exists and, since

$$u(M_u) = 1, \quad \text{one has} \qquad \sum_{L \subseteq N,\ a \in L\ \&\ u(L) > 0} u(L) \geqslant 2.$$

The condition (A.2) implies $\sum_{R \in \mathcal{T}_u,\ a \in R} u(R) \leqslant -2$. As $u(T) = -1$ for $T \in \mathcal{T}_u$ (by Corollary 1) it means $a \in W_u$. □

Lemma 6 (*about positive and negative sets*). *If $0 \neq u \in \mathbb{Z}^{\mathcal{P}(N)}$ satisfies (A)-(C) and $L \in \mathcal{L}_u^{max}$ then $|\{T \in \mathcal{T}_u;\ L \subseteq T\}|$ is either $u(L)$ or $u(L) + 1$. In particular, given $L \in \mathcal{L}_u$, there exists $T \in \mathcal{T}_u^{max}$ with $L \subset T$. Moreover, the following "sandwich" property is valid:*

$$\forall S, T \in \mathcal{T}_u \quad \textit{if } S \subset T \textit{ then } \exists K \in \mathcal{L}_u\ S \subset K \subset T.$$

Proof. First, we show $|\{T \in \mathcal{T}_u;\ L \subseteq T\}| \geqslant u(L)$. As $L \in \mathcal{L}_u^{max}$ and $u(M_u) = 1$, write using Corollary 1 and (C):

$$u(L) + 1 - |\{T \in \mathcal{T}_u; L \subseteq T\}| = \sum_{K, L \subseteq K \& u(K) > 0} u(K) + \sum_{K, L \subseteq K \& u(K) < 0} u(K) = \sum_{K, L \subseteq K} u(K) \leqslant 1.$$

Hence, $u(L) + 1 - |\{T \in \mathcal{T}_u;\ L \subseteq T\}| \leqslant 1$ gives the desired inequality.

Second, we show $|\{T \in \mathcal{T}_u;\ L \subseteq T\}| \leqslant u(L) + 1$. By the condition (4), which follows from (B), write:

$$u(L) + 1 - |\{T \in \mathcal{T}_u; L \subseteq T\}| = \sum_{K, L \subseteq K} u(K) \geqslant 0,$$

which gives the required inequality.

The "sandwich" property follows easily from (†). It says

$$\sum_{K, S \subseteq K \subseteq T \ \&\ u(K)>0} u(K) + \underbrace{\sum\nolimits_{K \in \mathcal{T}_u, S \subseteq K \subseteq T} u(K)}_{\leqslant -2} = \sum_{K, S \subseteq K \subseteq T} u(K) \geqslant -1,$$

implying the existence of K with $S \subseteq K \subseteq T$ and $u(K) > 0$. Of course, $S \neq K \neq T$. Thus, $S \subset K \subset T$ allows us to deduce $K \in \mathcal{L}_u$. □

Definition 3 (*final negative set*).
Given a non-zero imset $u \in \mathbb{Z}^{\mathcal{P}(N)}$ satisfying (A)–(C), a negative set $T \in \mathcal{T}_u$ will be called *final* (in u) if

$$T \setminus W_u \neq \emptyset \quad \& \quad T \cap W_u \in \mathcal{L}_u \cup \{\emptyset\}.$$

The class of final negative sets in u will be denoted by $\mathcal{T}_u^{fin}$. If a negative set $T \in \mathcal{T}_u$ satisfies $T \cap W_u = \emptyset$ it is called *isolated*.

Corollary 2. *If $0 \neq u \in \mathbb{Z}^{\mathcal{P}(N)}$ satisfies (A)–(C) then $\mathcal{T}_u^{fin} \subseteq \mathcal{T}_u^{max}$. Moreover, every isolated negative set is final. In particular, if $W_u = \emptyset$ then $\mathcal{T}_u^{fin} = \mathcal{T}_u$.*

Proof. If $T \in \mathcal{T}_u^{fin}$ then consider $a \in T \setminus W_u$. By the definition of the center W_u, the only set from $\mathcal{T}_u$ covering a is T. Thus, there is no strict superset of T in $\mathcal{T}_u$.

As any $T \in \mathcal{T}_u$ is non-empty (by Corollary 1) the second claim is evident. □

Now, we are able to describe how Conjecture 2 can perhaps be verified. One can use the procedure from [17, Section 7] for testing whether a given imset is standard. It is a recursive procedure, based on successive decomposition of the tested imset imset u over N. The reduction step allows one to transform the original task to the task to test another imset w over a smaller set $M \subset N$. However, to perform that step one has to find a final negative set T in sense of Definition 3. Note that such set always exists for a standard imset u over N; however, we omit the proof in this paper.

Thus, to verify Conjecture 2 by induction on $|N|$, one has to show the following:

- to prove that any (adapted imset) $u \in \mathbb{Z}^{\mathcal{P}(N)}$ satisfying (A)–(C) has a final negative set T,
- to observe that the imset $\tilde{w} = u - \delta_N + \delta_T + \delta_M - \delta_{T \cap M}$, where $M = N \setminus (T \setminus W_u)$, vanishes outside $\mathcal{P}(M)$,
- to observe that (these observations imply) that $\sum_{S \subseteq N} |u(S)| \leqslant 2\ |N|$ holds any imset u satisfying (A)–(C),
- to show that the restriction w of $\tilde{w}$ to M satisfies (A)–(C).

This is basically what has to be done because the other tests in the procedure from [17, Section 7] are passed by $u \in \mathbb{Z}^{\mathcal{P}(N)}$ satisfying (A)-(C), as shown in this appendix.

References

[1] S.A. Andersson, D. Madigan, M.D. Perlman, A characterization of Markov equivalence classes for acyclic digraphs, The Annals of Statistics 25 (1997) 505–541.
[2] E.A. Bender, J.R. Goldman, On the applications of Möbius inversion in combinatorial analysis, The American Mathematical Monthly 82 (1975) 789–803.
[3] R.R. Bouckaert, Bayesian belief networks: from construction to evidence, Ph.D Thesis, University of Utrecht, 1995.
[4] R.R. Bouckaert, R. Hemmecke, S. Lindner, M. Studený, Efficient algorithms for conditional independence inference, Journal of Machine Learning Research, accepted for publication.
[5] D.M. Chickering, Optimal structure identification with greedy search, Journal of Machine Learning Research 3 (2002) 507–554.
[6] L.M. de Campos, J.G. Castellano, Bayesian network learning algorithms using structural restrictions, International Journal of Approximate Reasoning 74 (2007) 233–254.
[7] M. Grabish, Set functions over finite sets: transformations and integrals, in: E. Pap (Ed.), Handbook of Measure Theory II in Chapter 34, Elsevier, 2002, pp. 1381–1402.
[8] R. Hemmecke, J. Morton, A. Shiu, B. Sturmfels, O. Wienand, Three counter-examples on semigraphoids, Combinatorics Probability and Computing 17 (2008) 239–257.
[9] S.L. Lauritzen, Graphical Models, Clarendon Press, 1996.
[10] R.E. Neapolitan, Learning Bayesian Networks, Pearson Prentice Hall, 2004.
[11] A. Schrijver, Theory of Linear and Integer Programming, John Wiley, 1986.
[12] G. Schwarz, Estimation the dimension of a model, The Annals of Statistics 6 (1978) 461–464.
[13] T. Silander, T. Roos, P. Myllymäki, Learning locally minimax optimal Bayesian networks, International Journal of Approximate Reasoning 51 (2010) 544–557.
[14] M. Studený, R.R. Bouckaert, T. Kočka, Extreme supermodular set functions over five variables, Research Report No. 1977, Institute of Information Theory and Automation, Prague, January 2000.
[15] M. Studený, Probabilistic Conditional Independence Structures, Springer Verlag, 2005.

[16] M. Studený, Mathematical aspects of learning Bayesian networks: Bayesian quality criteria, Research Report No. 2234, Institute of Information Theory and Automation, Prague, December 2008.
[17] M. Studený, J. Vomlel, A reconstruction algorithm for the essential graph, International Journal of Approximate Reasoning 50 (2009) 385–413.
[18] M. Studený, J. Vomlel, R. Hemmecke, A geometric view on learning Bayesian network structures, International Journal of Approximate Reasoning 51 (2010) 578–586.
[19] M. Studený, R. Hemmecke, S. Lindner, Characteristic imset: a simple algebraic representative of a Bayesian network structure, in: P. Myllymäki, T. Roos, T. Jaakkola (Eds.), Proceedings of PGM 2010, HIIT Publications, 2010, pp. 257–264. www.helsinki.fi/PGM2010/proceedings.html.