On polyhedral approximations of polytopes for learning Bayes nets

Milan Studený* David Haws†

July 26, 2011

Abstract

We review three vector encodings of Bayesian network structures. The first one has recently been applied by Jaakkola et al. [4], the other two use special integral vectors formerly introduced, called *imsets* [11, 13]. The central topic is the comparison of outer polyhedral approximations of the corresponding polytopes. We show how to transform the inequalities suggested by Jaakkola et al. to the framework of imsets. The result of our comparison is the observation that the implicit polyhedral approximation of the standard imset polytope suggested in [14] gives a closer approximation than the (transformed) explicit polyhedral approximation from [4]. Finally, we confirm a conjecture from [14] that the above-mentioned implicit polyhedral approximation of the standard imset polytope is an LP relaxation of the polytope.

1 Introduction

Bayesian networks (BNs) are popular graphical statistical models widely used both in probabilistic reasoning [8] and statistics [5]. They are attributed to acyclic directed graphs whose nodes correspond to the variables in consideration. The motivation for this report is learning the BN structure [7] from data by maximizing a quality (= scoring) criterion. The criterion is a real function of a BN structure (= of a graph) and of a database; its value says how much the BN structure given by the graph is good to explain the occurrence of the database.

However, different (acyclic directed) graphs can define the same statistical model, in which case the graphs are *Markov equivalent*. Thus, a usual requirement on the criterion is that it should be *score equivalent*, which means, it ascribes the same value to equivalent graphs. Another traditional technical requirement is that the criterion should be *decomposable* – for details see [2].

*Institute of Information Theory and Automation of the ASCR, studeny@utia.cas.cz
†University of Kentucky, Dept. of Statistics, dchaws@gmail.com

Since the aim is learning the BN structure (= statistical model) some researchers prefer to have a unique representative for every BN structure and to understand the criterion as a function of such unique representatives. A traditional unique graphical representative of the BN structure is the *essential graph* of the corresponding Markov equivalence class of acyclic directed graphs, which is a special graph allowing both directed and undirected edges – for details see [1].

The basic idea of an algebraic approach to learning, proposed in connection with conditional independence structures [11], is to represent every BN structure by a certain integral vector (= a vector with integers as components), called the *standard imset*. This is also a unique BN representative. The advantage of this algebraic approach is that every score equivalent and decomposable criterion becomes an affine function of the standard imset.

It has been shown in [12] that the standard imsets are vertices of a certain polytope, called the *standard imset polytope*. This allows one to re-formulate the learning task as a *linear programming* (LP) problem. However, to apply standard LP methods one needs the polyhedral description of the polytope. In [14], a conjecture about an implicit polyhedral characterization of the standard imset polytope has been presented. The weaker version of the conjecture was that the polyhedron given by those inequalities is an LP relaxation of the polytope.

Suitable transformation of an LP problem often simplifies things. Therefore, in [13], an alternative algebraic representative for the BN structure, called the *characteristic imset*, has been introduced. It is obtained from the standard imset by an invertible affine transformation; however, unlike the standard imset, the characteristic imset is always a zero-one vector. This opens the way to the application of advanced methods of *integer programming* (IP) in this area. Nonetheless, the crucial question of polyhedral characterization of the (transformed) polytope remain to be answered.

Jaakkola et al. [4] have also proposed to apply the methods of linear and integer programming to learning BN structures. They have used a straightforward zero-one encoding of acyclic directed graphs and transformed the task of maximizing the quality criterion to an IP problem. The main difference is that their vector codes are not unique BN representatives. On the other hand, they provide an explicit polyhedral LP relaxation of their polytope, which allows one to use the methods of IP.

In this report, we transform the inequalities suggested by Jaakkola et al. to the framework of imsets. First, we show that the implicit polyhedral approximation of the standard imset polytope suggested in [14] gives a closer approximation than the (transformed) explicit polyhedral approximation from [4]. Second, we show that the transformed inequalities give an explicit LP relaxation of the standard/characteristic imset polytope. A consequence of this fact is the proof of the weaker version of the conjecture from [14].

2 Notation and terminology

Throughout the paper N is a finite set of *variables* which has least two elements: $|N| \geq 2$. Its power set, denoted by $\mathcal{P}(N)$, is the class of its subsets $\{A;\ A \subseteq N\}$. For any $\ell = 1, 2$, we use a special notation

$$\mathcal{P}_\ell(N) \equiv \{A \subseteq N;\ |A| \geq \ell\}$$

for the class of subsets of N of cardinality at least ℓ. The symbol $U \subset V$ will mean $U \subseteq V$, $U \neq V$.

We deal with directed graphs (without loops) having N as the set of nodes and call them *directed graphs over* N. Such a graph is specified by a collection of arrows $j \to i$, where $i, j \in N$, $i \neq j$; the set $pa_G(i) \equiv \{j \in N;\ j \to i\}$ is (called) the set of *parents* of node $i \in N$. A directed cycle in G is a sequence of nodes $i_1, \ldots, i_n$, $n \geq 3$ such that $i_r \to i_{r+1}$ in G for $r = 1, \ldots, n-1$ and $i_n = i_1$. A directed graph is *acyclic* if it has no directed cycle. A well-known equivalent definition is that there exists an ordering $i_1, \ldots i_{|N|}$ of nodes of G *consistent with the direction of arrows* in G, which means $i_r \to i_s$ in G implies $r < s$. Clearly, every acyclic directed graph G has at least one *initial node*, that is, a node i with $pa_G(i) = \emptyset$.

We also deal with real vectors, elements of $\mathbb{R}^M$, where M is a non-empty finite set. By *lattice points* in $\mathbb{R}^M$ we mean integral vectors, that is, vectors whose components are integers (= elements of $\mathbb{Z}^M$). In this paper, M has additional structure; typically, it is $\mathcal{P}(N)$ or $\mathcal{P}_2(N)$, in which cases the lattice points are called *imsets*. To write formulas for imsets we will use the following notation: given $A \subseteq N$, the corresponding *basic vector* will be denoted by δ_A:

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

A special *semi-elementary imset* $\mathsf{u}_{\langle A,B|C\rangle}$ is associated with any (ordered) triplet of pairwise disjoint sets $A, B, C \subseteq N$:

$$\mathsf{u}_{\langle A,B|C\rangle} \equiv \delta_C - \delta_{A\cup C} - \delta_{B\cup C} + \delta_{A\cup B\cup C}\,,$$

which, in the context of [11], encodes the corresponding conditional independence statement $A \perp\!\!\!\perp B\,|\,C$. The imsets will be denoted using sans serif fonts, e.g. u or c; general vectors by bold lower-case letters, e.g. $\boldsymbol{b}$ or $\boldsymbol{\eta}$. They are interpreted as column vectors.

Matrices will be denoted by bold capitals, e.g. $\boldsymbol{A}$ or $\boldsymbol{C}$. The symbol $\boldsymbol{A}^\top$ denotes the transpose of $\boldsymbol{A}$. An invertible matrix $\boldsymbol{A}$ is *unimodular* if it is integral (= has integers as entries) and its determinant is $+1$ or -1 (see § 4.1 in [9]); an equivalent definition is that both $\boldsymbol{A}$ and its inverse $\boldsymbol{A}^{-1}$ are integral, that is, the mappings $\boldsymbol{b} \mapsto \boldsymbol{A}\boldsymbol{b}$ and $\boldsymbol{c} \mapsto \boldsymbol{A}^{-1}\boldsymbol{c}$ ascribe lattice points to lattice points.

By a full row rank matrix we mean an $m \times n$-matrix which has m linearly independent columns (= has rank m). The concept of unimodularity was extended in § 19.1 of [9] to matrices of this kind. A full row rank $m \times n$ matrix $\boldsymbol{A}$

is *unimodular* if every $m \times m$-submatrix has determinant $+1$, -1 or 0; equivalently, if any of its invertible $m \times m$-submatrix $\boldsymbol{B}$ is unimodular. A matrix $\boldsymbol{A}$ is *totally unimodular* if any of its (square) submatrix has determinant $+1$, 0 or -1.

We also deal with special classes of subsets of N. More specifically, we will consider non-empty classes $\mathcal{A}$ of non-empty subsets of N which are *closed under supersets*. These are classes $\emptyset \neq \mathcal{A} \subseteq \mathcal{P}_1(N)$ satisfying

$$S \in \mathcal{A}\,, \quad S \subseteq T \subseteq N \quad \Rightarrow \quad T \in \mathcal{A}\,.$$

Every such class $\mathcal{A}$ is characterized by the class $\mathcal{A}_{\min}$ of its *mimimal sets* with respect to inclusion:

$$\mathcal{A}_{\min} \equiv \{ S \in \mathcal{A};\; \forall\, T \subset S \quad T \notin \mathcal{A} \}\,.$$

Of course, $\mathcal{I} = \mathcal{A}_{\min}$ is a non-empty subclass of $\mathcal{P}_1(N)$ consisting of *incomparable sets*, which means

$$\forall\, S, T \in \mathcal{I}, \qquad S \neq T \quad \Rightarrow \quad [\, S \setminus T \neq \emptyset \;\;\&\;\; T \setminus S \neq \emptyset \,]\,.$$

Conversely, given a non-empty class $\mathcal{I} \subseteq \mathcal{P}_1(N)$ of incomparable sets the corresponding class $\mathcal{A}$ closed under supersets satisfying $\mathcal{I} = \mathcal{A}_{\min}$ is as follows:

$$\mathcal{A} = \{ S \subseteq N;\; \exists\, T \in \mathcal{I} \quad T \subseteq S \}\,.$$

Finally, in the proofs, we sometimes use Dirac's delta-symbol to shorten the notation. Specifically, the notation $\delta(\star\star)$, where $\star\star$ is a predicate (= statement), means a zero-one function whose value is $+1$ if the statement $\star\star$ is valid and whose value is 0 if the statement $\star\star$ does not hold.

3 Three ways of encoding Bayes nets

3.1 Straightforward zero-one encoding of a directed graph

Jaakkola et al. [4] used a special method for vector encoding (acyclic) directed graphs over N. Their 0-1-vectors $\boldsymbol{\eta}$ have components indexed by pairs $(i|B)$, where $i \in N$ and $B \subseteq N \setminus \{i\}$. Although their intention was to encode acyclic directed graphs only, one can formally encode any directed graph in this way. Specifically, given a directed graph G over N, the vector $\boldsymbol{\eta}_G$ encoding G is defined as follows:

$$\eta_G(i|B) = 1 \quad \Leftrightarrow \quad B = pa_G(i), \qquad \eta_G(i|B) = 0 \;\; \text{otherwise.}$$

Example 1 Consider $N = \{a, b, c\}$ and $G : a \leftrightarrows b \leftarrow c$. It is a directed graph, but not an acyclic one. We have $pa_G(a) = \{b\}$, $pa_G(b) = \{a, c\}$, $pa_G(c) = \emptyset$. Thus, $\eta_G(a|\{b\}) = 1$, $\eta_G(b|\{a, c\}) = 1$, $\eta_G(c|\emptyset) = 1$, and $\eta_G(i|B) = 0$ otherwise.

The polytope studied by Jaakkola et. al. [4] is defined as the convex hull of the set of vectors $\boldsymbol{\eta}_G$, where G runs over all acyclic directed graphs over N.

3.1.1 Jaakkola et al.'s polyhedral approximation

The (outer) polyhedral approximation J of the above polytope proposed in [4] is given by the following constraints:

- "simple" *non-negativity constraints*:

$$\eta(i|B) \geq 0 \quad \text{for every } i \in N,\, B \subseteq N \setminus \{i\} \tag{1}$$

  ($|N| \cdot 2^{|N|-1}$ inequality constraints),

- *equality constraints*:

$$\sum_{B \subseteq N \setminus \{j\}} \eta(j|B) = 1 \quad \text{for all } j \in N \tag{2}$$

  ($|N|$ equality constraints),

- *cluster inequalities*, which correspond to sets $C \subseteq N$, $|C| \geq 2$:

$$1 \leq \sum_{i \in C} \sum_{B \subseteq N \setminus \{i\},\, B \cap C = \emptyset} \eta(i|B) \equiv \sum_{i \in C} \sum_{D \subseteq N \setminus C} \eta(i|D) \tag{3}$$

  ($2^{|N|} - |N| - 1$ cluster inequalities).

Taking into account the equality constraints (2) for $i \in C$, (3) takes the form

$$1 \leq \sum_{i \in C} \left[1 - \sum_{B \subseteq N \setminus \{i\},\, B \cap C \neq \emptyset} \eta(i|B) \right].$$

Remark No cluster inequality for $C = \emptyset$ is defined; the cluster inequalities for $|C| = 1$ are omitted because they follow trivially from the equality constraints.

Example 2 In case $N = \{a, b, c\}$ every $\boldsymbol{\eta}$-vector has length 12 and its components decompose into three blocks that correspond to variables a, b and c. Thus, one has twelve non-negativity constraints, three equality constraints and four cluster inequalities of two types:

- $1 \leq \eta(a|\emptyset) + \eta(a|\{c\}) + \eta(b|\emptyset) + \eta(b|\{c\})$, (for $C = \{a, b\}$)
- $1 \leq \eta(a|\emptyset) + \eta(b|\emptyset) + \eta(c|\emptyset)$. (for $C = \{a, b, c\}$)

The constraints (1) and (2) are clearly valid for any vector $\boldsymbol{\eta}_G$ of a directed graph G; the inequalities (3) hold in the acyclic case – see Lemma 4.

3.1.2 Jaakkola et al.'s approximation is an LP relaxation

The polyhedral approximation from § 3.1.1 is an *LP relaxation* of the corresponding polytope, by which we mean that the only lattice points in the approximation are the lattice points in the polytope. First, we observe that the polyhedron J' given by non-negativity and equality constraints is an integral polytope.

Lemma 3 *Let J' be the polyhedron given by (1) and (2). Then J' is a polytope whose vertices are just the codes of (general) directed graphs over N. Moreover, the only lattice points in J' are its vertices.*

Proof. Let $\boldsymbol{\eta}$ belong to J'. For every block of components of $\boldsymbol{\eta}$ corresponding to $i \in N$, the constraints define a vector in a "probability simplex". Assuming $\boldsymbol{\eta}$ is a vertex of J', for each $i \in N$, the respective block has to be a vertex of that simplex, that is, a 0-1-vector having just one component 1. If $B(i)$ is the set indexing such a component for $i \in N$, we get the corresponding graph G with $\boldsymbol{\eta} = \boldsymbol{\eta}_G$ by drawing arrows from the elements of $B(i)$ to i, for every $i \in N$. Clearly, this defines a one-to-correspondence between (general) directed graphs over N and vertices of J'.

Let $\boldsymbol{\eta}$ be a lattice point in J'. Within the block given by $i \in N$, components are non-negative integers. Thus, if one of them exceeds 1, the sum exceeds 1. Hence, $\boldsymbol{\eta}$ is a 0-1-vector. At most one component in a block is 1 since otherwise the sum exceeds 1, and at least one is 1 since otherwise the sum is 0. $\square$

Lemma 4 *Let J be the polyhedron given by constraints (1)-(3). Then the lattice points in J are exactly the codes of acyclic directed graphs over N.*

Proof. Every lattice point in J is a lattice point in J', and, therefore, by Lemma 3, encodes a (uniquely determined) directed graph G.

Consider the cluster equality (3) for $C \subseteq N$, $|C| \geq 2$ and the vector $\boldsymbol{\eta}_G$ (encoding a directed graph G). For every $i \in C$, the $\eta_G(i|D)$ term is typically 0 and only once 1, namely in the case $D = pa_G(i)$. Thus, the inner expression for i in (3), namely $\sum_{D \subseteq N \setminus C} \eta_G(i|D)$ is either 0 or 1. The latter happens if and only if $pa_G(i) \cap C = \emptyset$. That means, the cluster inequality for C says there exists at least one $i \in C$ with $pa_G(i) \cap C = \emptyset$. Of course, this is true if G is acyclic.

Now, we are going to show the converse: the cluster inequalities for $\boldsymbol{\eta}_G$ imply that G is acyclic. We start with applying the cluster inequality for $C = N$ and find $i_1 \in N$ with $pa_G(i_1) = \emptyset$. Thus, i_1 is an initial node in G and we fix it. If $|N \setminus \{i_1\}| \geq 2$ we take $C = N \setminus \{i_1\}$ and apply the cluster inequality for it. It says there exists $i_2 \in C = N \setminus \{i_1\}$ with $pa_G(i_2) \cap C = \emptyset$, that is, $pa_G(i_2) \subseteq \{i_1\}$ ($\equiv i_2$ is the initial node in the induced subgraph $G_{N \setminus \{i_1\}}$).

Again, if $|N \setminus \{i_1, i_2\}| \geq 2$ we continue with $C = N \setminus \{i_1, i_2\}$, and so on. In this way, we find iteratively an ordering $i_1, \dots i_{|N|}$ consistent with the direction of arrows in G. This already implies G is acyclic. $\square$

3.2 Standard imsets

Standard imsets introduced [11] have components indexed by subsets $T \subseteq N$. Given an acyclic directed graph G over N, the *standard imset* u_G encoding G is defined as follows:

$$\mathsf{u}_G = \delta_N - \delta_\emptyset + \sum_{i \in N} \left[\delta_{pa_G(i)} - \delta_{\{i\} \cup pa_G(i)} \right].$$

A basic property of standard imsets is that they are unique representatives of Bayesian network structures. This means, one has $\mathsf{u}_G = \mathsf{u}_H$ if and only if G and H are independence equivalent acyclic directed graphs (= define the same Bayesian network structure) – see Corollary 7.1 in [11]. In [12], it was proposed to study the *standard imset polytope*, defined as the convex hull of the set of vectors u_G, where G runs over all acyclic directed graphs with N vertices.

3.2.1 Outer approximation of the standard imset polytope

In [14], an outer approximation of the standard imset polytope in terms of linear constraints was suggested. More specifically, three types of constraints were considered (for $\mathsf{u} = \mathsf{u}_G$):

- *equality contraints*:

$$\sum_{T \subseteq N} \mathsf{u}(T) = 0, \qquad \forall\, j \in N \sum_{T \subseteq N,\, j \in T} \mathsf{u}(T) = 0\,, \tag{4}$$

  which implies that u-vectors are determined uniquely by their components $\mathsf{u}(T)$ for $T \subseteq N$, $|T| \geq 2$,

- *specific inequality contraints* of the form:

$$\sum_{T \in \mathcal{A}} \mathsf{u}(T) \leq 1\,, \tag{5}$$

  where $\mathcal{A}$ is a non-empty class of non-empty subsets of N, closed under supersets,

- *non-specific inequality contraints* of the form:

$$\langle m, \mathsf{u} \rangle \equiv \sum_{T \subseteq N} m(T) \cdot \mathsf{u}(T) \geq 0\,, \tag{6}$$

  where m is a (representative on an extreme standardized) supermodular function. Here, by a *supermodular function* is meant a real function m on the power set $\mathcal{P}(N)$ ($\equiv$ a vector in $\mathbb{R}^{\mathcal{P}(N)}$) such that

$$m(E \cup F) + m(E \cap F) \geq m(E) + m(F) \quad \text{for every } E, F \subseteq N.$$

  It is *standardized* if $m(T) = 0$ whenever $|T| \leq 1$.

Note that the class of standardized supermodular functions on $\mathcal{P}(N)$ is a pointed rational polyhedral cone, and, therefore, has finitely many extreme rays. Each extreme ray contains a uniquely determined non-zero lattice point whose components have no common prime divisor (this is the representative of the extreme ray). Therefore, (6) gives in fact finitely many linear inequality constraints on $\mathsf{u} = \mathsf{u}_G$. The problem is that one has to compute those representatives of extreme supermodular functions, which is a difficult computational task. The representatives were computed for $|N| \leq 5$ [10].

Thus, in comparison with the polyhedral approximation (of the $\boldsymbol{\eta}$-polytope) mentioned in § 3.1.1, this polyhedral approximation (of the standard imset polytope) is implicit. This is a disadvantage from the practical point of view because to apply common methods of linear programming one still needs to explicate the considered inequality constraints for any $|N|$.

Example 5 In case $N = \{a, b, c\}$ every u-vector has the length 8. There are four equality constraints (4) which break into two types:

- $\mathsf{u}(\emptyset) = -\mathsf{u}(a) - \mathsf{u}(b) - \mathsf{u}(c) - \mathsf{u}(\{a, b\}) - \mathsf{u}(\{a, c\}) - \mathsf{u}(\{b, c\}) - \mathsf{u}(\{a, b, c\})$,
- $\mathsf{u}(a) = -\mathsf{u}(\{a, b\}) - \mathsf{u}(\{a, c\}) - \mathsf{u}(\{a, b, c\})$. (for $j = a$)

Therefore, the dimension (of the standard imset polytope) is 4 and the u-vectors are determined by their components for sets $\{a, b\}$, $\{a, c\}$, $\{b, c\}$ and $\{a, b, c\}$.

As concerns specific inequality constraints, every non-empty class of $\mathcal{A}$ of non-empty subsets of N closed under supersets is uniquely determined by the class $\mathcal{A}_{\min}$ of its minimal sets with respect to inclusion. One has eighteen such classes which break into eight types. For example, $\mathcal{A}_{\min} = \{ab, ac, bc\}$ gives the inequality

$$\mathsf{u}(\{a, b\}) + \mathsf{u}(\{a, c\}) + \mathsf{u}(\{b, c\}) + \mathsf{u}(\{a, b, c\}) \leq 1 \,.$$

As concerns non-specific inequality constraints, the cone of standardized supermodular functions has five extreme rays in case $|N| = 3$ [10], which leads to five inequalities breaking into three types:

- $\mathsf{u}(\{a, b, c\}) \geq 0$,
- $\mathsf{u}(\{a, b\}) + \mathsf{u}(\{a, b, c\}) \geq 0$,
- $\mathsf{u}(\{a, b\}) + \mathsf{u}(\{a, c\}) + \mathsf{u}(\{b, c\}) + 2 \cdot \mathsf{u}(\{a, b, c\}) \geq 0$.

Note that the described system of inequalities can be reduced; some of the specific inequalities appear to follow from the non-specific ones in combination with equality constraints and other specific inequalities. For example, if $\mathcal{A}_{\min}$ consists of one singleton only, then the respective specific inequality (5) is vacuous because it trivially follows from the equality constrains (4). Actually, all specific inequalities with $\mathcal{A}_{\min}$ containing a singleton are superfluous in case $|N| = 3$. However, this is not true in case $|N| \geq 4$.

The constraints (4)-(6) were conjectured in [14] to completely characterize the standard imset polytope and this conjecture was verified for $|N| \leq 4$. Nevertheless, one perhaps does not need a complete facet description (= polyhedral characterization) of the polytope. To apply some advanced methods of integer programming the confirmation of a weaker version of the conjecture might be enough. The weaker version of the conjecture from [14] is that the polyhedron given by (4)-(6) is an LP relaxation of the standard imset polytope.

Before writing this report, we confirmed computationally the weaker version for $|N| = 5$. The extreme rays of the cone of supermodular functions for $|N| = 5$ were obtained from [10] and independently computed using *4ti2* [17], thus giving the non-specific inequality constraints (6). Specific inequality constraints (5) were obtained from [14], where it was also calculated that there are 8,782 standard imsets for $|N| = 5$. Since the characteristic imsets (described in § 3.3) are 0-1-vectors and are in one-to-one correspondence to the standard imsets, we simply enumerated all vectors in $\{0,1\}^{\mathcal{P}_2(N)}$, applied the inverse transform (11) to get the corresponding u-vectors, and tested whether they satisfied the above inequalities. By operating over $\mathcal{P}_2(N)$, and properly modifying the above inequalities, the equality constraints (4) were satisfied. We verified that there were exactly 8,782 integer solutions to the constraints (4)-(6) for $|N| = 5$.

3.2.2 $\boldsymbol{\eta}$ to standard imset

Taking into account the definition of $\boldsymbol{\eta}_G$, it is easy to see that u_G is obtained from $\boldsymbol{\eta}_G$ by applying the following mapping $\boldsymbol{\eta} \mapsto \mathsf{u}^{\boldsymbol{\eta}}$. For any $T \subseteq N$, we put

$$\mathsf{u}^{\boldsymbol{\eta}}(T) = \delta_N(T) - \delta_\emptyset(T) + \sum_{i \in N} \sum_{B \subseteq N \setminus \{i\}} \eta(i|B) \cdot \{\delta_B(T) - \delta_{\{i\} \cup B}(T)\}\,. \qquad (7)$$

This is clearly an affine mapping, ascribing lattice points to lattice points. Assuming $\boldsymbol{\eta}$ belongs to the linear subspace specified by equality constraints (2), we re-write (7) as follows:

$$\begin{aligned}
\mathsf{u}^{\boldsymbol{\eta}}(T) &= \delta_N(T) - \delta_\emptyset(T) \\
&+ \sum_{i \in N} \eta(i|\emptyset) \cdot \{\delta_\emptyset(T) - \delta_{\{i\}}(T)\} + \sum_{i \in N} \sum_{\emptyset \neq B \subseteq N \setminus \{i\}} \eta(i|B) \cdot \{\delta_B(T) - \delta_{\{i\} \cup B}(T)\} \\
&\overset{(2)}{=} \delta_N(T) - \delta_\emptyset(T) + \sum_{i \in N} \{1 - \sum_{\emptyset \neq B \subseteq N \setminus \{i\}} \eta(i|B)\} \cdot \{\delta_\emptyset(T) - \delta_{\{i\}}(T)\} + \ldots \\
&= \underbrace{\delta_N(T) + (|N| - 1) \cdot \delta_\emptyset(T) - \sum_{i \in N} \delta_{\{i\}}(T)}_{\mathsf{u}^\emptyset(T) \in \mathbb{Z}} \\
&- \sum_{i \in N} \sum_{\emptyset \neq B \subseteq N \setminus \{i\}} \eta(i|B) \cdot \underbrace{\{\delta_\emptyset(T) - \delta_{\{i\}}(T) - \delta_B(T) + \delta_{\{i\} \cup B}(T)\}}_{\mathsf{u}_{\langle i,B|\emptyset\rangle}(T) \in \{-1,0,+1\}}\,,
\end{aligned}$$

where $\mathsf{u}^\emptyset$ denotes the standard imset corresponding to the empty graph over N and $\mathsf{u}_{\langle i,B|\emptyset\rangle}$ the semi-elementary imset encoding $i \perp\!\!\!\perp B \,|\, \emptyset$.

Briefly, if $\boldsymbol{\eta}$ satisfies (2) then

$$\mathsf{u}^{\boldsymbol{\eta}} = \mathsf{u}^{\emptyset} - \sum_{i\in N} \sum_{\emptyset\neq B\subseteq N\setminus\{i\}} \eta(i|B) \cdot \mathsf{u}_{\langle i,B|\emptyset\rangle}.$$

In particular, $\mathsf{u} = \mathsf{u}^{\boldsymbol{\eta}}$ belongs to the linear subspace specified by equality constraints (4). This is because these equalities hold for both $\mathsf{u}^{\emptyset}$ and any $\mathsf{u}_{\langle i,B|\emptyset\rangle}$. Note that the converse is true as well (we leave an easy proof to the reader): if u satisfies (4) then there exists $\boldsymbol{\eta}$ satisfying (2) such that $\mathsf{u} = \mathsf{u}^{\boldsymbol{\eta}}$. In particular, (4) is the exact translation of (2) into the framework of standard imsets.

3.3 Characteristic imsets

The characteristic imset (for an acyclic directed graph G), introduced in [13], is obtained from the standard imset by an affine transformation. More specifically, first, the *portrait* p_G of the standard imset u_G is obtained by a linear transform; second, the portrait is subtracted from the constant 1-vector and the *characteristic imset* c_G is obtained:

$$\begin{aligned} \mathsf{p}(S) &= \sum_{T,\, S\subseteq T\subseteq N} \mathsf{u}(T) \qquad \text{for } S\subseteq N, && (8)\\ \mathsf{c}(S) &= 1-\mathsf{p}(S) \qquad \text{for } S\subseteq N. && (9) \end{aligned}$$

Clearly, the equality constraints (4) are translated into the following tacit restrictions on c-vectors:

$$\mathsf{c}(S) = 1 \qquad \text{for } S\subseteq N,\ |S|\leq 1\,. \tag{10}$$

Therefore, for an acyclic directed graph G over N, the components of the characteristic imset c_G for $|S|\leq 1$ are ignored and c_G is formally considered to be an element of $\mathbb{Z}^{\mathcal{P}_2(N)}$.

The mapping $\mathsf{u}\mapsto\mathsf{c}$ determined by (8)-(9) is invertible: one can compute back the standard imset by the formula

$$\mathsf{u}(T) = \sum_{S,\, T\subseteq S\subseteq N} (-1)^{|S\setminus T|} \cdot \underbrace{[\,1-\mathsf{c}(S)\,]}_{\mathsf{p}(S)} \qquad \text{for } T\subseteq N. \tag{11}$$

Indeed, to see it fix $S\subseteq N$, substitute (11) (with S replaced by D) into the expression for the portrait $\mathsf{p}(S)$ and change the order of summation:

$$\begin{aligned} \sum_{T,\, S\subseteq T\subseteq N} \mathsf{u}(T) &= \sum_{T,\, S\subseteq T\subseteq N} \ \sum_{D,\, T\subseteq D\subseteq N} (-1)^{|D\setminus T|}\cdot \mathsf{p}(D)\\ &= \sum_{D,\, S\subseteq D\subseteq N} \mathsf{p}(D)\cdot \underbrace{\sum_{T,\, S\subseteq T\subseteq D} (-1)^{|D\setminus T|}}_{\delta_S(D)} = \mathsf{p}(S)\,. \end{aligned}$$

Since the transformation is one-to-one, two acyclic directed graph G and H are independence equivalent if and only if $\mathsf{c}_G = \mathsf{c}_H$. Thus, the characteristic imset is also a unique Bayesian network structure representative.

3.3.1 Advantage of characteristic imsets

Since standard and characteristic imsets are in one-to-one correspondence, one can transform the inequality constraints from § 3.2.1 into the framework of characteristic imsets – see § 4.1.3 and § 4.2 for further details. One important consequence of these transformed constraints are *basic inequalities for characteristic imsets* valid in the acyclic case:

Corollary 6 *The constraints (4)-(6) on* u *imply the inequalities* $0 \leq \mathsf{c}(S) \leq 1$, $S \subseteq N$ *for the imset* c *ascribed to* u *by (8)-(9).*

Proof. Because of (9), we show $0 \leq \mathsf{p}(S) \leq 1$ for $S \subseteq N$. First, (4) says $\mathsf{p}(S) = 0$ for $|S| \leq 1$. Given $S \subseteq N$, $|S| \geq 2$ the class of sets $\mathcal{A} = \{T;\ S \subseteq T \subseteq N\}$ is closed under supersets and, by (5), $\mathsf{p}(S) \leq 1$. On the other hand, in (6), among the (representatives of extreme) supermodular functions we find the function

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

In particular, among the non-specific inequality constraints is the inequality $\mathsf{p}(S) = \sum_{T,\, S \subseteq T} \mathsf{u}(T) \equiv \langle m^{S\uparrow}, \mathsf{u} \rangle \geq 0$. □

In particular, every characteristic imset c_G (for an acyclic directed graph G) is a 0-1-vector, which is a fact emphasized already in [13], which is important from the point of view of (possible future application of) methods of integer programming.

Another advantage of characteristic imsets is that they are closer to the graphical description (of Bayesian network structures) than standard imsets. Specifically, for $S \subseteq N$, $|S| \geq 2$ one has

$$\mathsf{c}_G(S) = 1 \quad \Leftrightarrow \text{ there exists } i \in S \text{ with } S \setminus \{i\} \subseteq pa_G(i), \tag{12}$$

and there exists a polynomial algorithm for transforming the characteristic imset c_G into the respective *essential graph*, which is a traditional unique graphical representative of the Bayesian network structure given by G – see [13].

3.3.2 $\boldsymbol{\eta}$ to characteristic imset

Lemma 7 *The characteristic imset* c_G *is a linear function of* $\boldsymbol{\eta}_G$ *given by*

$$\mathsf{c}(S) = \sum_{i \in S} \sum_{B,\, S \setminus \{i\} \subseteq B \subseteq N \setminus \{i\}} \eta(i|B) \qquad \textit{where } |S| \geq 1. \tag{13}$$

Proof. Given $S \subseteq N$, substitute (7) into (8) and change the order of summation:

$$\begin{aligned}
\mathsf{p}(S) &= \sum_{T,\, S\subseteq T\subseteq N} \big[\, \delta_N(T) - \delta_\emptyset(T) + \sum_{i\in N} \sum_{B\subseteq N\setminus\{i\}} \eta(i|B)\cdot\{\delta_B(T) - \delta_{\{i\}\cup B}(T)\}\,\big] \\
&= \sum_{T,\, S\subseteq T\subseteq N} \delta_N(T) - \sum_{T,\, S\subseteq T\subseteq N} \delta_\emptyset(T) \\
&\quad + \sum_{i\in N} \sum_{B\subseteq N\setminus\{i\}} \eta(i|B)\cdot\big\{ \sum_{T,\, S\subseteq T\subseteq N} \delta_B(T) - \sum_{T,\, S\subseteq T\subseteq N} \delta_{\{i\}\cup B}(T)\,\big\} \\
&= 1 - \delta_\emptyset(S) + \sum_{i\in N} \sum_{B\subseteq N\setminus\{i\}} \eta(i|B)\cdot\big\{\, \delta(S\subseteq B) - \delta(S\subseteq\{i\}\cup B)\,\big\}\,.
\end{aligned}$$

Realize that the expression $\delta(S\subseteq B) - \delta(S\subseteq\{i\}\cup B)$ vanishes if either $S\subseteq B$ or $S\setminus(\{i\}\cup B)\neq\emptyset$, otherwise it is -1. Thus, assuming $|S|\geq 1$, one has

$$\begin{aligned}
\mathsf{p}(S) &= 1 + \sum_{i\in N} \sum_{B\subseteq N\setminus\{i\}} \eta(i|B)\cdot(-1)\cdot\delta(i\in S,\ S\subseteq\{i\}\cup B) \\
&= 1 - \sum_{i\in S} \sum_{B\subseteq N\setminus\{i\}} \eta(i|B)\cdot\delta(S\setminus\{i\}\subseteq B)\,,
\end{aligned}$$

because, in case $i\in S$, then $S\subseteq\{i\}\cup B$ is equivalent to $S\setminus\{i\}\subseteq B$. Taking (9) into consideration we get (13). □

Let us call the mapping given by (13) the *characteristic transformation*. It can formally be applied to any $\boldsymbol{\eta}$-vector, in particular, to the code $\boldsymbol{\eta}_G$ of a general directed graph G. Thus, we get a formula for the "quasi-characteristic" imset (= an element of $\mathbb{Z}^{\mathcal{P}_2(N)}$) ascribed to a graph over N:

$$\mathsf{c}_G(S) = \text{number of super-terminal nodes in } S \quad \text{for } S\subseteq N,\ |S|\geq 2. \tag{14}$$

Here, a *super-terminal node* (in S) means $i\in S$ such that for all $j\in S\setminus\{i\}$ one has $j\to i$ in G. Indeed, having fixed S, $|S|\geq 2$ and $i\in S$, the expression $\sum_{B,\, S\setminus\{i\}\subseteq B\subseteq N\setminus\{i\}} \eta_G(i|B)$ is either 0 or 1 depending upon $S\setminus\{i\}\subseteq pa_G(i)$. Observe that (12) is a special case (14) since, in case of an acyclic directed graph, any set S has at most one super-terminal node.

Example 8 Consider the graph G from Example 1. Then $\mathsf{c}_G(\{a,c\}) = 0$, $\mathsf{c}_G(\{b,c\}) = \mathsf{c}_G(\{a,b,c\}) = 1$ and $\mathsf{c}_G(\{a,b\}) = 2$. Observe that c_G does not satisfy the basic constrains $0\leq\mathsf{c}\leq 1$ valid in acyclic case. This is because G is not acyclic.

4 Transformation of inequality constraints

In § 3.2.2 and § 3.3.2, we have described mappings which transform the $\boldsymbol{\eta}$-vectors used by Jaakkola et al. [4] to standard/characteristic imsets. The advantage of the $\boldsymbol{\eta}$-polytope is the existence of a good (= explicit) outer polyhedral approximation (see Lemma 4 in § 3.1.1). In this section, we characterize the image of that polyhedral approximation (by the above maps) and compare the transformed approximation (of $\boldsymbol{\eta}$-polytope) with the approximation of the standard imset polytope from § 3.2.1. The main technical difficulty we have to tackle is that the mappings transforming $\boldsymbol{\eta}$-vectors to imsets are many-to-one. Another feature is that the transformation raises the number of linear constraints. To clarify the reasons for that, in § 4.1 we first deal with the transformation of elementary constraints (1)-(2) and, later, in § 4.2, with the transformation of cluster inequalities (3).

4.1 Transformation of elementary $\boldsymbol{\eta}$-constraints

Now, the question of our interest is to transform the constraints (1)-(2) only, that is, to characterize the form of the inequalities of the image of the polyhedron J' from Lemma 3. Let us start with an example, illustrating our method.

Example 9 Consider $N = \{a, b, c\}$, the polyhedron J' and the characteristic transformation $\boldsymbol{\eta} \mapsto \mathsf{c}$ given by (13). The idea is to transform each vertex of J' and take the convex hull R of the images of vertices. Because of linearity of the map $\boldsymbol{\eta} \mapsto \mathsf{c}$, the polytope R is the image of J'. Thus, it is enough to find the facet description of R; this is the exact translation of (1)-(2) then.

The vertices of J' are exactly the codes of general directed graphs (see Lemma 3) and their images are given by (14). Thus, the (permutation type representatives of) images of vertices of J' were obtained in this way. Here they are (the order of component is ab, ac, bc, abc):

$$[0,0,0,0], [1,0,0,0], [2,0,0,0], [2,1,0,0], [1,1,0,0], [1,1,1,0],$$
$$[1,1,0,1], [2,1,0,1], [2,2,0,1], [1,1,1,1], [2,1,1,1],$$
$$[2,1,1,2], [2,2,1,2], [2,2,2,3].$$

Remaining images can be obtained by permutation of first 3 components. We computed the facet-description of their convex hull R by *Polymake* [3]. The result had fifteen inequalities. Here, we only recorded the (permutation) types of obtained inequalities:

- $0 \le \mathsf{c}(ab)$,
- $0 \le 2 - \mathsf{c}(ab)$,
- $0 \le 3 - \mathsf{c}(ab) - \mathsf{c}(ac) - \mathsf{c}(bc) + \mathsf{c}(abc)$,
- $0 \le \mathsf{c}(abc)$,
- $0 \le 1 + \mathsf{c}(ab) - \mathsf{c}(abc)$,
- $0 \le \mathsf{c}(ab) + \mathsf{c}(ac) - \mathsf{c}(abc)$,

- $0 \leq \mathsf{c}(ab) + \mathsf{c}(ac) + \mathsf{c}(bc) - 2\mathsf{c}(abc)$.

To make sure we computed the vertices of the polyhedron given by these inequalities. The (permutation) type representatives are as follows:

$[0,0,0,0], [2,0,0,0], [2,1,0,0], [1,1,0,1], [2,1,0,1], [2,2,0,1], [2,1,1,2], [2,2,2,3]$.

We observe that some of images of vertices of J' are convex combinations of the others: for example, $[1,0,0,0]$ comes from $[0,0,0,0]$ and $[2,0,0,0]$. Note that the original polyhedron J' was given by twelve inequalities (and three equality constraints). Since R is given by fifteen inequality (and four implicit equality) constraints, the transformation to the framework of characteristic imsets raised the number of inequality constraints.

Another interesting observation is that the obtained fifteen inequalities in fact coincide with the translation of specific inequality constraints (5) to the framework of characteristic imsets in case $N = \{a,b,c\}$ – see Example 14 for details.

This leads to a natural conjecture that Jaakkola et al.'s elementary constraints (1)-(2) are equivalent to our specific constraints for any $|N|$. We confirm this conjecture below, directly by considering the transformation of $\boldsymbol{\eta} \mapsto \mathsf{u}$. Later, we transform the specific constraints to the framework of characteristic imsets (see § 4.1.3).

4.1.1 Translation to the framework of standard imsets

Thus, the task is to characterize in terms of u the image (by $\boldsymbol{\eta} \mapsto \mathsf{u}^{\boldsymbol{\eta}}$) of the polytope J' given by non-negativity and equality constraints. More specifically, we wish to have a finite system of linear inequalities on u which together with (4) – see § 3.2.2 – characterize those $\mathsf{u} \in \mathbb{R}^{\mathcal{P}(N)}$ for which

$$\exists\, \boldsymbol{\eta} \;\text{ satisfying (1),(2) and }\; \mathsf{u}^{\boldsymbol{\eta}}(T) = \mathsf{u}(T) \;\text{ for any } T \subseteq N, |T| \geq 2\,. \tag{15}$$

This task can equivalently be formulated as follows. Let us put $m \equiv 2^{|N|} - 1$, $n \equiv |N| \cdot 2^{|N|-1}$ and consider a special $m \times n$ matrix $\boldsymbol{A}$, whose

- rows correspond to sets $T \subseteq N$, $|T| \geq 1$,
- columns correspond to pairs $(i|B)$ where $i \in N$, $B \subseteq N \setminus \{i\}$.

More specifically, the entry $\boldsymbol{a}\,[\,T,(i|B)\,]$ of $\boldsymbol{A}$ is given by

$$\begin{array}{rcll} \boldsymbol{a}\,[\,T,(i|B)\,] & = & \delta_{\{i\}\cup B}(T) - \delta_B(T) & \text{if } \;|T| \geq 2\,, \\ \boldsymbol{a}\,[\,T,(i|B)\,] & = & \delta_{\{i\}}(T) & \text{if } \;|T| = 1\,. \end{array} \tag{16}$$

Moreover, to any $\mathsf{u} \in \mathbb{R}^{\mathcal{P}_2(N)}$, we ascribe a column m-vector $\boldsymbol{b}_{\mathsf{u}}$ whose components $b_{\mathsf{u}}[\,T\,]$ are specified as follows:

$$\begin{array}{rcll} b_{\mathsf{u}}[\,T\,] &=& \delta_N(T) - \mathsf{u}(T) & \text{if } \ |T| \geq 2\,, \\ b_{\mathsf{u}}[\,T\,] &=& 1 & \text{if } \ |T| = 1\,. \end{array}$$

Then (15) is equivalent to the condition

$$\exists\, \boldsymbol{\eta} \in \mathbb{R}^n \ \ \text{satisfying } \boldsymbol{\eta} \geq 0 \text{ and } \boldsymbol{A}\boldsymbol{\eta} = \boldsymbol{b}_{\mathsf{u}}. \tag{17}$$

Indeed, (1) means $\boldsymbol{\eta} \geq 0$, while (2) for $j \in N$ is the requirement that the component of $\boldsymbol{b}_{\mathsf{u}}$ for $T = \{j\}$, which is 1, coincides with the respective component of $\boldsymbol{A}\boldsymbol{\eta}$:

$$1 = \sum_{(i|B)} \boldsymbol{a}\,[\,T, (i|B)\,] \cdot \eta(i|B) = \sum_{i \in N} \sum_{B \subseteq N \setminus \{i\}} \delta_{\{i\}}(\{j\}) \cdot \eta(i|B) = \sum_{B \subseteq N \setminus \{j\}} \eta(j|B)\,.$$

Analogously, for fixed $T \subseteq N$, $|T| \geq 2$, $\mathsf{u}(T) = \mathsf{u}^{\boldsymbol{\eta}}(T)$ has, by (7), the form

$$\mathsf{u}(T) = \delta_N(T) - \sum_{i \in N} \sum_{B \subseteq N \setminus \{i\}} \underbrace{\{\delta_{\{i\} \cup B}(T) - \delta_B(T)\}}_{\boldsymbol{a}\,[\,T, (i|B)\,]} \cdot \eta(i|B)$$

and can be expressed equivalently as

$$\sum_{i \in N} \sum_{B \subseteq N \setminus \{i\}} \boldsymbol{a}\,[\,T, (i|B)\,] \cdot \eta(i|B) = \delta_N(T) - \mathsf{u}(T) \equiv b_{\mathsf{u}}(T)\,,$$

which means the components of $\boldsymbol{A}\boldsymbol{\eta}$ and $\boldsymbol{b}_{\mathsf{u}}$ for T coincide.

Now, Farkas' lemma (see Corollary 7.1d in [9]) applied to $\boldsymbol{A}$ and $\boldsymbol{b}_{\mathsf{u}}$ says that (17) is equivalent to the requirement:

$$\forall\, y \in \mathbb{R}^m \quad \boldsymbol{A}^{\top} y \geq 0 \quad \Rightarrow \quad \boldsymbol{b}_{\mathsf{u}}^{\top} y \geq 0\,. \tag{18}$$

To simplify this requirement we re-write the condition $\boldsymbol{A}^{\top} y \geq 0$ in this form:

$$\forall\, i \in N \qquad y(\{i\}) \geq 0, \tag{19}$$

$$\forall\, S \subseteq N, |S| = 2, \ \forall\, i \in S \qquad y(S) + y(\{i\}) \geq 0, \tag{20}$$

$$\forall\, S \subseteq N, |S| \geq 3, \ \forall\, i \in S \qquad y(S) + y(\{i\}) - y(S \setminus \{i\}) \geq 0. \tag{21}$$

Indeed, the rows of $\boldsymbol{A}^{\top}$ correspond to pairs $(i|B)$, $i \in N$, $B \subseteq N \setminus \{i\}$. If $i \in N$ and $B = \emptyset$ then the component of $\boldsymbol{A}^{\top} y$ for $(i|\emptyset)$ is as follows:

$$\sum_{\emptyset \neq T \subseteq N} \boldsymbol{a}\,[\,T, (i|\emptyset)\,] \cdot y(T) = \sum_{|T|=1} \delta_{\{i\}}(T) \cdot y(T) = y(\{i\})\,,$$

because $\boldsymbol{a}\,[\,T, (i|\emptyset)\,] = 0$ for $|T| \geq 2$. This gives (19). If $i \in N$, $B \subseteq N \setminus \{i\}$ with $|B| = 1$, then $\boldsymbol{a}\,[\,T, (i|B)\,] = \delta_{\{i\} \cup B}(T)$ for $|T| \geq 2$ and one can write

$$\begin{aligned} &\sum_{\emptyset \neq T \subseteq N} \boldsymbol{a}\,[\,T, (i|B)\,] \cdot y(T) \\ &= \sum_{|T|=1} \delta_{\{i\}}(T) \cdot y(T) + \sum_{|T| \geq 2} \delta_{\{i\} \cup B}(T) \cdot y(T) = y(\{i\}) + y(\{i\} \cup B)\,, \end{aligned}$$

which leads to (20) for $S = \{i\} \cup B$. Finally, if $i \in N$, $B \subseteq N \setminus \{i\}$ with $|B| \geq 2$ then

$$\begin{aligned}
&\sum_{\emptyset \neq T \subseteq N} \boldsymbol{a}\,[T,(i|B)] \cdot y(T) \\
&= \sum_{|T|=1} \delta_{\{i\}}(T) \cdot y(T) + \sum_{|T|\geq 2} \{\delta_{\{i\}\cup B}(T) - \delta_B(T)\} \cdot y(T) \\
&= y(\{i\}) + y(\{i\} \cup B) - y(B)\,,
\end{aligned}$$

which leads to (21) for $S = \{i\} \cup B$.

The next step is to show that $\{y \in \mathbb{R}^m; \boldsymbol{A}^\top y \geq 0\}$ is a pointed (rational polyhedral) cone and characterize its extreme rays. In fact, we show that the rays correspond to non-empty classes of sets $\mathcal{A} \subseteq \mathcal{P}_1(N)$ closed under supersets. More specifically, we ascribe a vector $y_{\mathcal{A}} \in \mathbb{R}^m$ to any such class $\mathcal{A}$ by:

$$y_{\mathcal{A}}(T) \equiv \delta(T \in \mathcal{A}) - |\{j \in N;\ \{j\} \in \mathcal{A} \;\&\; \{j\} \subset T\}| \qquad \text{for } T \in \mathcal{P}_1(N)\,. \tag{22}$$

Here is the crucial observation:

Lemma 10 *A vector $y \in \mathbb{R}^m$ satisfies (19)-(21) if and only if it is a conic combination (= a linear combination with non-negative real coefficients) of vectors $y_{\mathcal{A}}$ for classes $\emptyset \neq \mathcal{A} \subseteq \mathcal{P}_1(N)$ closed under supersets.*

Proof. First, we leave to the reader to verify that any such vector $y_{\mathcal{A}}$ satisfies (19)-(21), which implies the sufficiency of the condition.

To verify the converse implication, we ascribe to any $y \in \mathbb{R}^m$ satisfying (19)-(21) the class of sets

$$\mathcal{A}_y = \{S \in \mathcal{P}_1(N);\ \exists\, T \in \mathcal{P}_1(N),\ T \subseteq S \;\; y(T) \neq 0\}\,,$$

which is clearly closed under supersets and non-empty if $y \neq 0$. The idea is to prove the converse implication by induction on $|\mathcal{A}_y|$. If $|\mathcal{A}_y| = 0$ then $y \equiv 0$ and the claim that y is a conic combination of those vectors is evident. If $|\mathcal{A}_y| \geq 1$ then it is enough to find some $\beta > 0$ such that $y' \equiv y - \beta \cdot y_{\mathcal{A}}$ satisfies (19)-(21) and $|\mathcal{A}_{y'}| < |\mathcal{A}_y|$.

Since now we fix $y \in \mathbb{R}^m$, $y \neq 0$ satisfying (19)-(21) and put:

$$\mathcal{A} \equiv \mathcal{A}_y, \quad y_* \equiv y_{\mathcal{A}}, \quad Y \equiv \{i \in N;\ y(\{i\}) \neq 0\}\,.$$

Observe a few basic facts:

$$y(S) = 0 \quad \text{for } S \in \mathcal{P}_1(N) \setminus \mathcal{A}, \qquad y(S) > 0 \quad \text{for } S \in \mathcal{A}_{\min}\,.$$

Indeed, assuming $S \in \mathcal{A}_{\min}$ one has $y(S) \neq 0$. If $|S| = 1$ then (19) implies $y(S) > 0$. If $|S| = 2$ then $\{i\} \notin \mathcal{A}$ for both $i \in S$. Hence, $y(\{i\}) = 0$ and (20) gives $y(S) > 0$. If $|S| \geq 3$ and $i \in S$, then both $\{i\} \notin \mathcal{A}$ and $S \setminus \{i\} \notin \mathcal{A}$ and (21) gives $y(S) > 0$.

In particular, since $\{j\} \in \mathcal{A}_{\min}$ for $j \in Y$, and $\{i\} \notin \mathcal{A}$ for $i \notin Y$,

$$\beta \equiv \min\{y(T);\ T \in \mathcal{A}_{\min}\} > 0, \text{ and} \qquad y(\{j\}) \geq \beta > 0 \quad \text{for } j \in Y, \tag{23}$$
$$y(\{i\}) = 0 \quad \text{for } i \in N \setminus Y\,. \tag{24}$$

Further, we observe that y is non-decreasing set function on subsets of $N \setminus Y$. *Indeed, it is enough to show $T \subseteq S \subseteq N \setminus Y$, $|S \setminus T| = 1 \;\Rightarrow\; y(S) \geq y(T)$. Take $S \setminus T = \{i\}$; then $i \notin Y$ and $y(\{i\}) = 0$. If $|S| = 2$ then $T = \{j\}$ with $j \notin Y$ and $y(S) = y(S) + y(\{i\}) \geq 0 = y(T)$ follows from (20) and (24). If $|S| \geq 3$ then (21) says $y(S) + 0 - y(T) \geq 0$.*

This implies:

$$S \in \mathcal{A}\,,\ S \cap Y = \emptyset \quad \Rightarrow \quad y(S) \geq \beta\,. \tag{25}$$

Indeed, it is enough to find $T \in \mathcal{A}_{\min}$, $T \subseteq S$ (of course, $T \cap Y = \emptyset$) and combine $y(S) \geq y(T)$ with $y(T) \geq \beta$, which follows from the definition of β in (23).

Finally, also have:

$$S \in \mathcal{P}_1(N),\ |S \cap Y| \leq 1 \quad \Rightarrow \quad y(S) \geq 0\,. \tag{26}$$

Indeed, this was verified in cases $|S| = 1$ and $|S \cap Y| = 0$ in (23)-(25). Assume $|S| \geq 2$ and $|S \cap Y| = 1$ and use the induction on $|S|$. If $|S| = 2$ then $S = \{i, j\}$ with $i \notin Y$ and $j \in Y$ and (20)+(24) give $y(S) \geq -y(\{i\}) = 0$. If $|S| \geq 3$ then choose $i \in S \setminus Y$ and write by (21)+(24) $y(S) \geq y(S \setminus \{i\}) - y(\{i\}) = y(S \setminus \{i\})$. Now, $y(S \setminus \{i\}) \geq 0$ follows from the induction premise.

To smooth later considerations let us gather the observations about $y_* = y_{\mathcal{A}}$ defined in (22). For singletons we have:

$$y_*(\{i\}) = 1 \quad \text{for } i \in Y, \qquad y_*(\{i\}) = 0 \quad \text{for } i \notin Y.$$

Given $S \subseteq N$, $|S| = 2$ we have:

$$\begin{aligned} y_*(S) &= 1 && \text{if } S \cap Y = \emptyset,\ S \in \mathcal{A}. \\ y_*(S) &= 0 && \text{if either } [\,S \cap Y = \emptyset \ \&\ S \notin \mathcal{A}\,] \text{ or } |S \cap Y| = 1. \\ y_*(S) &= -1 && \text{if } S \subseteq Y. \end{aligned}$$

For $S \subseteq N$, $|S| \geq 3$ we have:

$$\begin{aligned} y_*(S) &= 1 && \text{if } S \cap Y = \emptyset,\ S \in \mathcal{A}. \\ y_*(S) &= 0 && \text{if } S \cap Y = \emptyset,\ S \notin \mathcal{A}. \\ y_*(S) &= 1 - |S \cap Y| && \text{if } S \cap Y \neq \emptyset. \end{aligned}$$

To show that

$$y' \equiv y - \beta \cdot y_*$$

satisfies (19), that is, $y'(\{i\}) \geq 0$ for $i \in N$, we distinguish two cases.

- If $i \notin Y$ then $y_*(\{i\}) = 0$ and (19) for y implies the same equality for y'.

- If $i \in Y$ then $y'(\{i\}) = y(\{i\}) - \beta \cdot y_*(\{i\}) = y(\{i\}) - \beta \cdot 1 = y(\{i\}) - \beta \geq 0$ owing to (23).

To show that y' satisfies (20), that is, $y'(S) + y'(\{i\}) \geq 0$ for $S \subseteq N$, $|S| = 2$ and $i \in S$ we distinguish five cases.

- If $S \subseteq Y$ then $i \in Y$ and $y_*(S) + y_*(\{i\}) = (-1) + 1 = 0$ and (20) for y implies the same equality for y', no matter what β is.
- If $|S \cap Y| = 1$, $i \notin Y$ then $y_*(S) + y_*(\{i\}) = 0 + 0 = 0$ and (20) for y implies what is desired, for the same reason.
- If $|S \cap Y| = 1$, $i \in Y$ then $y'(S) + y'(\{i\}) = y(S) - \beta \cdot y_*(S) + y(\{i\}) - \beta \cdot y_*(\{i\}) = y(S) - \beta \cdot 0 + y(\{i\}) - \beta \cdot 1 = y(S) + y(\{i\}) - \beta$. However, $y(\{i\}) - \beta \geq 0$ by (23) and $y(S) \geq 0$ by (26), which implies what is desired.
- If $S \cap Y = \emptyset$, $S \notin \mathcal{A}$ then $i \notin Y$ and $y_*(S) + y_*(\{i\}) = 0 + 0 = 0$ and (20) for y implies what is desired,
- If $S \cap Y = \emptyset$, $S \in \mathcal{A}$ then $i \notin Y$ and by (24) $y'(S) + y'(\{i\}) = y(S) - \beta \cdot y_*(S) + y(\{i\}) - \beta \cdot y_*(\{i\}) = y(S) - \beta \cdot 1 + 0 - \beta \cdot 0 = y(S) - \beta$. The desired inequality follows from (25).

To show that y' satisfies (21), that is, $y'(S) + y'(\{i\}) - y'(S \setminus \{i\}) \geq 0$ for $S \subseteq N$, $|S| \geq 3$ and $i \in S$ we distinguish seven cases.

- If $S \cap Y = \emptyset$, $S \setminus \{i\} \in \mathcal{A}$ then $S \in \mathcal{A}$ and $i \notin Y$. Thus, $y_*(S) + y_*(\{i\}) - y_*(S \setminus \{i\}) = (+1) + 0 - (+1) = 0$ and (21) for y implies the same inequality for y'.
- If $S \cap Y = \emptyset$, $S \notin \mathcal{A}$ (which implies $S \setminus \{i\} \notin \mathcal{A}$) then $y_*(S) + y_*(\{i\}) - y_*(S \setminus \{i\}) = 0 + 0 - 0 = 0$ and (21) for y implies what is desired.
- If $S \cap Y = \emptyset$, $S \in \mathcal{A}$, $S \setminus \{i\} \notin \mathcal{A}$ then $y(\{i\}) = 0 = y(S \setminus \{i\})$ and we can write $y'(S) + y'(\{i\}) - y'(S \setminus \{i\}) = y(S) - \beta \cdot y_*(S) + y(\{i\}) - \beta \cdot y_*(\{i\}) - y(S \setminus \{i\}) + \beta \cdot y_*(S \setminus \{i\}) = y(S) - \beta \cdot 1 + 0 - \beta \cdot 0 - 0 + \beta \cdot 0 = y(S) - \beta$, which is non-negative by (25).
- If $S \cap Y \neq \emptyset$, $i \notin Y$ then $S \cap Y = (S \setminus \{i\}) \cap Y$ and $y_*(S) + y_*(\{i\}) - y_*(S \setminus \{i\}) = (+1 - |S \cap Y|) + 0 - (+1 - |(S \setminus \{i\}) \cap Y|) = 0$. Thus, (21) for y implies what is desired.
- If $S \cap Y \neq \emptyset$, $i \in Y$, $(S \setminus \{i\}) \cap Y \neq \emptyset$ then $|S \cap Y| = 1 + |(S \setminus \{i\}) \cap Y|$ and $y_*(S) + y_*(\{i\}) - y_*(S \setminus \{i\}) = (+1 - |S \cap Y|) + (+1) - (+1 - |(S \setminus \{i\}) \cap Y|) = 0$. Thus, (21) for y implies what is desired.
- If $S \cap Y \neq \emptyset$, $i \in Y$, $(S \setminus \{i\}) \cap Y = \emptyset$, $S \setminus \{i\} \in \mathcal{A}$ then $|S \cap Y| = 1$ and $y_*(S) + y_*(\{i\}) - y_*(S \setminus \{i\}) = (+1 - 1) + (+1) - (+1) = 0$ and (21) for y implies what is desired.

- If $S \cap Y \neq \emptyset$, $i \in Y$, $(S \setminus \{i\}) \cap Y = \emptyset$, $S \setminus \{i\} \notin \mathcal{A}$ then also $|S \cap Y| = 1$ and $y(S \setminus \{i\}) = 0$, which allows us to write $y'(S) + y'(\{i\}) - y'(S \setminus \{i\}) = y(S) - \beta \cdot y_*(S) + y(\{i\}) - \beta \cdot y_*(\{i\}) - y(S \setminus \{i\}) + \beta \cdot y_*(S \setminus \{i\}) = y(S) - \beta \cdot (1-1) + y(\{i\}) - \beta \cdot 1 - 0 + \beta \cdot 0 = y(S) + y(\{i\}) - \beta$. However, $y(\{i\}) - \beta \geq 0$ by (23) and $y(S) \geq 0$ by (26), which implies what is desired.

Thus, y' satisfies (19)-(21) and, because of the choice of β, $y'(T) = 0$ for at least one $T \in \mathcal{A}_{\min}$ and $|\mathcal{A}_{y'}| < |\mathcal{A}_y|$, which concludes the induction step. Indeed, realize that, by (22), $y_*(T) \equiv y_{\mathcal{A}}(T) = 0$ for $T \in \mathcal{P}_1(N) \setminus \mathcal{A}$. $\square$

Now, Lemma 10 allows us to re-formulate the requirement (18) in the form of finitely many conditions on u:

$$\forall \emptyset \neq \mathcal{A} \subseteq \mathcal{P}_1(N) \text{ closed under supersets} \quad \boldsymbol{b}_{\mathsf{u}}^{\top} y_{\mathcal{A}} \geq 0\,. \tag{27}$$

Indeed, if $y \in \mathbb{R}^m$ is such that $\boldsymbol{A}^{\top} y \geq 0$ and $y = \sum \lambda_{\mathcal{A}} \cdot y_{\mathcal{A}}$, $\lambda_{\mathcal{A}} \geq 0$, then $\boldsymbol{b}_{\mathsf{u}}^{\top} y = \sum \lambda_{\mathcal{A}} \cdot \boldsymbol{b}_{\mathsf{u}}^{\top} y_{\mathcal{A}} \geq 0$.

It remains to reformulate, given such an $\mathcal{A}$, the condition $\boldsymbol{b}_{\mathsf{u}}^{\top} y_{\mathcal{A}} \geq 0$. Assuming $|N| \geq 2$, denote for this purpose $A \equiv \{i \in N;\ \{i\} \in \mathcal{A}\}$ and write using the definition of $\boldsymbol{b}_{\mathsf{u}}$ and $y_{\mathcal{A}}$ from (22):

$$\begin{aligned} 0 \quad \leq \quad & \boldsymbol{b}_{\mathsf{u}}^{\top} y_{\mathcal{A}} = \sum_{|T|\geq 1} b_{\mathsf{u}}(T) \cdot y_{\mathcal{A}}(T) = \sum_{|T|=1} y_{\mathcal{A}}(T) + \sum_{|T|\geq 2} \{\delta_N(T) - \mathsf{u}(T)\} \cdot y_{\mathcal{A}}(T) \\ = \quad & \sum_{|T|=1} y_{\mathcal{A}}(T) + y_{\mathcal{A}}(N) - \sum_{|T|\geq 2} \mathsf{u}(T) \cdot y_{\mathcal{A}}(T) \\ = \quad & |A| + \underbrace{(1-|A|)}_{y_{\mathcal{A}}(N)} - \sum_{|T|\geq 2} \mathsf{u}(T) \cdot y_{\mathcal{A}}(T) = 1 - \sum_{|T|\geq 2} \mathsf{u}(T) \cdot y_{\mathcal{A}}(T)\,. \end{aligned}$$

Thus, $\boldsymbol{b}_{\mathsf{u}}^{\top} y_{\mathcal{A}} \geq 0$ is equivalent to $\sum_{|T|\geq 2} \mathsf{u}(T) \cdot y_{\mathcal{A}}(T) \leq 1$. To get even more elegant form of it, assume u satisfies (4) and observe

$$\begin{aligned} \sum_{|T|\geq 2} \mathsf{u}(T) \cdot |T \cap A| &= \sum_{|T|\geq 2} \mathsf{u}(T) \cdot \sum_{i \in A} \delta(i \in T) = \sum_{|T|\geq 2} \sum_{i \in A} \mathsf{u}(T) \cdot \delta(i \in T) \\ = \quad \sum_{i \in A} \sum_{|T|\geq 2} \mathsf{u}(T) \cdot \delta(i \in T) &= \sum_{i \in A} \sum_{|T|\geq 2,\, i \in T} \mathsf{u}(T) \overset{(4)}{=} \sum_{i \in A} -\mathsf{u}(\{i\})\,. \end{aligned}$$

Therefore, we can write by (22):

$$\begin{aligned} \sum_{|T|\geq 2} \mathsf{u}(T) \cdot y_{\mathcal{A}}(T) &= \sum_{|T|\geq 2} \mathsf{u}(T) \cdot \delta(T \in \mathcal{A}) - \sum_{|T|\geq 2} \mathsf{u}(T) \cdot |T \cap A| \\ = \quad \sum_{|T|\geq 2} \mathsf{u}(T) \cdot \delta(T \in \mathcal{A}) + \sum_{i \in A} \mathsf{u}(\{i\}) &= \sum_{|T|\geq 1} \mathsf{u}(T) \cdot \delta(T \in \mathcal{A}) = \sum_{T \in \mathcal{A}} \mathsf{u}(T)\,, \end{aligned}$$

which means that $\boldsymbol{b}_{\mathsf{u}}^{\top} y_{\mathcal{A}} \geq 0$ is equivalent to $\sum_{T \in \mathcal{A}} \mathsf{u}(T) \leq 1$. Thus, under validity of (4), (27) is equivalent to (5) and we have:

Corollary 11 *Provided $|N| \geq 2$, the condition (15) for $\mathsf{u} \in \mathbb{R}^{\mathcal{P}(N)}$ is equivalent to the simultaneous validity of (4) and (5).*

4.1.2 Remarks on the matrix $\boldsymbol{A}$

Consider again the $m\times n$ matrix $\boldsymbol{A}$ defined in (16); recall that $m=2^{|N|}-1$ and $n=|N|\cdot 2^{|N|-1}$. We have observed in § 4.1.1 that $\boldsymbol{A}$ plays a central role in the transition from $\boldsymbol{\eta}$-vectors to standard imsets. Now, we show that $\boldsymbol{A}$ has full row rank by deriving its its Hermite normal form (see § 4.1 in [9] for this concept).

Proposition 12 *The matrix $\boldsymbol{A}$ has Hermite normal form $[\boldsymbol{I}\ \mathbf{0}]$, where $\boldsymbol{I}$ is the $m\times m$ identity matrix and $\mathbf{0}$ the $m\times(n-m)$ zero matrix.*

Proof. The columns of $\boldsymbol{A}$ are indexed by pairs $(i|B)$ and given by

$$
\begin{array}{ll}
\boldsymbol{A}_{(i|\emptyset)}=\delta_{\{i\}} & \text{for } i\in N,\\
\boldsymbol{A}_{(i|j)}=\delta_{\{i\}}+\delta_{\{i,j\}} & \text{for } i,j\in N,\ i\neq j\\
\boldsymbol{A}_{(i|B)}=\delta_{\{i\}}-\delta_{B}+\delta_{\{i\}\cup B} & \text{for } i\in N,\ B\subseteq N\setminus\{i\},\ |B|\geq 2.
\end{array}
$$

Thus, $\delta_{\{i\}}=\boldsymbol{A}_{(i|\emptyset)}$ and $\delta_{\{i,j\}}=\boldsymbol{A}_{(i|j)}-\boldsymbol{A}_{(i|\emptyset)}$. To show by induction on $|T|\geq 1$ that δ_T can be written as an integer combination of columns of $\boldsymbol{A}$, assume $|T|\geq 3$ and choose a pair $(i|B)$ with $T=\{i\}\cup B$, $|B|\geq 2$. Then

$$\delta_T=\delta_{\{i\}\cup B}=\boldsymbol{A}_{(i|B)}-\delta_{\{i\}}+\delta_B,$$

where, by the induction hypothesis, the terms $\delta_{\{i\}}$ and δ_B can be written as integer combination of the columns of $\boldsymbol{A}$.

Thus, using elementary columns operations, $\boldsymbol{A}$ can be transformed such that it contains all m elementary column vectors δ_T. Using additional column operations, all other columns can be zeroed out. Therefore, using elementary column operations, $\boldsymbol{A}$ can be transformed to the form $[\boldsymbol{I}\ \mathbf{0}]$. □

Before writing this report, we verified computationally that $\boldsymbol{A}$ is unimodular, strongly unimodular, strongly k-modular, however not totally unimodular for $3\leq|N|\leq 6$ using software written by Matthias Walther available at `https://github.com/xammy/unimodularity-test`. This led us to a hypothesis that $\boldsymbol{A}$ is unimodular for any $|N|$. In § 4.1.4, we confirm this hypothesis.

4.1.3 Translation to the framework of characteristic imsets

We observed in § 4.1.1 that Jaakkola et al.'s elementary constraints (1)-(2) are transformed into u-constraints as (4)-(5). Transforming (4)-(5) into c-constraints is a simpler task because of the one-to-one correspondence $\mathsf{u}\leftrightarrow\mathsf{c}$ (see § 3.3). We already know that (4) takes the form of tacit restrictions on c-vectors (10). As concerns the specific inequality constraints (5), we show below that every such inequality, for $\emptyset\neq\mathcal{A}\subseteq\mathcal{P}_1(N)$ closed under supersets, is transformed into the framework of c-vectors as follows:

$$0\leq\sum_{S\subseteq N}\kappa_{\mathcal{A}}(S)\cdot\mathsf{c}(S)\,, \tag{28}$$

where the coefficients $\kappa_{\mathcal{A}}(-)$ are given by

$$\kappa_{\mathcal{A}}(S) \equiv \sum_{T\in\mathcal{A},\, T\subseteq S} (-1)^{|S\setminus T|} \qquad \text{for } S\subseteq N\,. \tag{29}$$

However, the formula (29) is not suitable to compute the coefficients. It is more appropriate to introduce them equivalently in terms of the class $\mathcal{A}_{\min}\equiv\mathcal{I}$ of minimal sets in $\mathcal{A}$. More specifically, let us introduce the class $\mathcal{C}(\mathcal{I})$ of possible unions of sets from a non-empty class $\mathcal{I}\subseteq\mathcal{P}_1(N)$ of incomparable sets:

$$\mathcal{C}(\mathcal{I}) \equiv \{S\subseteq N;\ \exists\,\emptyset\neq\mathcal{K}\subseteq\mathcal{I} \quad \text{such that } S=\textstyle\bigcup_{T\in\mathcal{K}} T\}.$$

Then can can compute the coefficients $\kappa_{\mathcal{A}}(-)$ recursively as follows:

$$\begin{array}{lll} \kappa_{\mathcal{A}}(S)=0 & & \text{if } S\subseteq N,\ S\notin\mathcal{C}(\mathcal{I}),\\ \kappa_{\mathcal{A}}(S)=1-\sum_{T\in\mathcal{C}(\mathcal{I}),\, T\subset S} \kappa_{\mathcal{A}}(T) & & \text{for } S\in\mathcal{C}(\mathcal{I})\,. \end{array} \tag{30}$$

This implies that $\kappa_{\mathcal{A}}(S)=1$ for $S\in\mathcal{A}_{\min}=\mathcal{I}$ and that $\kappa_{\mathcal{A}}$ has the more zeros the smaller $|\mathcal{A}_{\min}|$ is. Therefore, in the framework of characteristic imsets, it is more convenient to ascribe the (transformed) specific inequality constraints directly to classes $\emptyset\neq\mathcal{I}\subseteq\mathcal{P}_1(N)$ of incomparable sets.

Lemma 13 *Let* u *and* c *be imsets related by (8)-(9) and* $\emptyset\neq\mathcal{A}\subseteq\mathcal{P}_1(N)$ *a class of sets closed under supersets. Then the inequality (5) corresponding to* $\mathcal{A}$ *has the form (28), where the coeficients* $\kappa_{\mathcal{A}}(-)$ *are given by (30).*

Proof. The first observation is that the coefficients given by (29) satisfy

$$\kappa_{\mathcal{A}}(S)=0 \ \text{ for } S\subseteq N,\ S\notin\mathcal{A}, \quad \text{and} \quad \sum_{S\subseteq N}\kappa_{\mathcal{A}}(S)=\sum_{S\in\mathcal{A}}\kappa_{\mathcal{A}}(S)=1\,. \tag{31}$$

To verify it realize that $\mathcal{A}\subseteq\mathcal{P}(N)$ is closed under supersets and write:

$$\begin{aligned} \sum_{S\in\mathcal{A}}\kappa_{\mathcal{A}}(S) &= \sum_{S\in\mathcal{A}}\ \sum_{T\in\mathcal{A},\, T\subseteq S} (-1)^{|S\setminus T|} = \sum_{T\in\mathcal{A}}\ \sum_{S\in\mathcal{A},\, T\subseteq S} (-1)^{|S\setminus T|}\\ &= \sum_{T\in\mathcal{A}}\ \sum_{S,\, T\subseteq S\subseteq N} (-1)^{|S\setminus T|} = \sum_{T\in\mathcal{A}} \delta_N(T) = 1\,. \end{aligned}$$

To see that (5) is transformed into (28) we substitute the inverse formula (11) into it and use the fact $\mathcal{A}$ is closed under supersets:

$$\begin{aligned} 1 \ \geq\ & \sum_{T\in\mathcal{A}} \mathsf{u}(T) = \sum_{T\in\mathcal{A}}\ \sum_{S,\, T\subseteq S\subseteq N} (-1)^{|S\setminus T|}\cdot\mathsf{p}(S)\\ =\ & \sum_{S\in\mathcal{A}}\ \sum_{T\in\mathcal{A},\, T\subseteq S} \mathsf{p}(S)\cdot(-1)^{|S\setminus T|} = \sum_{S\in\mathcal{A}} \mathsf{p}(S)\cdot \underbrace{\sum_{T\in\mathcal{A},\, T\subseteq S} (-1)^{|S\setminus T|}}_{\kappa_{\mathcal{A}}(S)}\,. \end{aligned}$$

Thus, substitute (31) in that inequality and get

$$0 \le 1 - \sum_{S\in\mathcal{A}} \mathsf{p}(S)\cdot\kappa_{\mathcal{A}}(S) \stackrel{(31)}{=} \sum_{S\in\mathcal{A}} \kappa_{\mathcal{A}}(S) - \sum_{S\in\mathcal{A}} \kappa_{\mathcal{A}}(S)\cdot\mathsf{p}(S) = \sum_{S\in\mathcal{A}} \kappa_{\mathcal{A}}(S)\cdot\underbrace{[1-\mathsf{p}(S)]}_{\mathsf{c}(S)},$$

which is, owing to (9) and (31), nothing but (28).

It remains to show that (29) takes the form (30). An auxiliary fact is

$$\forall\, S\in\mathcal{A} \qquad \sum_{T\subseteq S} \kappa_{\mathcal{A}}(T) = 1\,. \tag{32}$$

Indeed, to see it, consider the class $\mathcal{A}_S \equiv \{T\subseteq S;\ T\in\mathcal{A}\}$, which is a class of subsets of S, closed under supersets. Moreoever, for any $T\in\mathcal{A}_S$, one has $\kappa_{\mathcal{A}}(T)=\kappa_{\mathcal{A}_S}(T)$, which implies by (31) applied to $\mathcal{A}_S$ and S in place of N that

$$1 = \sum_{T\in\mathcal{A}_S} \kappa_{\mathcal{A}_S}(T) = \sum_{T\in\mathcal{A},\, T\subseteq S} \kappa_{\mathcal{A}}(T) = \sum_{T\subseteq S} \kappa_{\mathcal{A}}(T)\,.$$

In the rest of the proof we write $\mathcal{I}$ in place of $\mathcal{A}_{\min}$ and omit the index in $\kappa_{\mathcal{A}}(-)$ and write $\kappa(-)$ only. For every $\mathcal{K}\subseteq\mathcal{I}$ we introduce the class of sets whose only subsets in $\mathcal{I}$ are elements of $\mathcal{K}$:

$$\mathcal{B}_{\mathcal{K}} \equiv \{S\subseteq N;\ K\subseteq S \text{ for } K\in\mathcal{K} \ \ \& \ \ L\setminus S\neq\emptyset \text{ for } L\in\mathcal{I}\setminus\mathcal{K}\ \}.$$

Of course, it may happen that $\mathcal{B}_{\mathcal{K}}$ is empty for some $\mathcal{K}\subseteq\mathcal{I}$. Nevertheless, the collection of classes $\mathcal{B}_{\mathcal{K}}$, where $\mathcal{K}$ runs over subsets of $\mathcal{I}$, form a partition of $\mathcal{P}(N)$. Moreover, every non-empty class $\mathcal{B}_{\mathcal{K}}$ has the least set (in sense of inclusion), namely $S_{\mathcal{K}}\equiv\bigcup_{T\in\mathcal{K}} T$. Observe that $\mathcal{K}=\emptyset$ leads to a non-empty class $\mathcal{B}_{\emptyset}=\mathcal{P}(N)\setminus\mathcal{A}$ with $S_{\emptyset}=\emptyset$. Since $\mathcal{I}$ consists of incomparable sets, every $S\in\mathcal{I}$ belongs to just one $\mathcal{B}_{\mathcal{K}}$ with $|\mathcal{K}|=1$, namely $\mathcal{K}=\{S\}$. The class $\mathcal{C}(\mathcal{I})$ defined above (30) then coincides with $\{S_{\mathcal{K}};\ \emptyset\neq\mathcal{K}\subseteq\mathcal{I}\ \text{ with } \mathcal{B}_{\mathcal{K}}\neq\emptyset\,\}$.

An easy consequence of (29) is that $\kappa(S)=0$ for $S\in\mathcal{B}_{\emptyset}$ ($= S\notin\mathcal{A}$) and $\kappa(S)=1$ for $S\in\mathcal{I}$. To verify (30) it is enough to show by induction on $|\mathcal{K}|$ the following two statements:

(i) $\kappa(S)=0$ for $S\in\mathcal{B}_{\mathcal{K}}$, $S\neq S_{\mathcal{K}}$,

(ii) $\forall\,|\mathcal{K}|\geq 1$ with $\mathcal{B}_{\mathcal{K}}\neq\emptyset \quad 1=\sum_{\mathcal{L}\subseteq\mathcal{K},\,\mathcal{B}_{\mathcal{L}}\neq\emptyset}\kappa(S_{\mathcal{L}})$.

Indeed, this is because for $\mathcal{L},\mathcal{K}\subseteq\mathcal{I}$ with $\mathcal{B}_{\mathcal{L}}\neq\emptyset\neq\mathcal{B}_{\mathcal{K}}$ one has $\mathcal{L}\subseteq\mathcal{K}$ if and only if $S_{\mathcal{L}}\subseteq S_{\mathcal{K}}$. We already know this is true in case $|\mathcal{K}|=0$. Now assume $|\mathcal{K}|\geq 1$ and the statements hold for any $\mathcal{L}\subset\mathcal{K}$. Consider arbitrary $S\in\mathcal{B}_{\mathcal{K}}$ and write using (32) and the fact that subsets of S must belong to $\mathcal{B}_{\mathcal{L}}$ for $\mathcal{L}\subseteq\mathcal{K}$:

$$\begin{aligned} 1 \ &\stackrel{(32)}{=}\ \sum_{T\subseteq S}\kappa(T) = \sum_{\mathcal{L}\subseteq\mathcal{K},\,\mathcal{B}_{\mathcal{L}}\neq\emptyset}\ \sum_{T\subseteq S,\,T\in\mathcal{B}_{\mathcal{L}}}\kappa(T) \\ &=\ \sum_{T\subseteq S,\,T\in\mathcal{B}_{\mathcal{K}}}\kappa(T) + \sum_{\mathcal{L}\subset\mathcal{K},\,\mathcal{B}_{\mathcal{L}}\neq\emptyset}\ \sum_{T\subseteq S,\,T\in\mathcal{B}_{\mathcal{L}}}\kappa(T)\,. \end{aligned} \tag{33}$$

Now, observe that the induction premise (i) applied to any $\mathcal{L} \subset \mathcal{K}$, $\mathcal{B}_{\mathcal{L}} \neq \emptyset$ says that κ vanishes in $\mathcal{B}_{\mathcal{L}}$ except for $S_{\mathcal{L}}$. In particular, for any $\mathcal{D} \subseteq \mathcal{B}_{\mathcal{L}}$ with $S_{\mathcal{L}} \in \mathcal{D}$ one has $\sum_{T\in\mathcal{D}} \kappa(T) = \kappa(S_{\mathcal{L}})$. This implies that the second term in (33) is $\sum_{\mathcal{L}\subset\mathcal{K},\, \mathcal{B}_{\mathcal{L}}\neq\emptyset} \kappa(S_{\mathcal{L}})$ and we have observed that

$$\forall\, S \in \mathcal{B}_{\mathcal{K}} \qquad \sum_{T\subseteq S,\, T\in\mathcal{B}_{\mathcal{K}}} \kappa(T) = 1 - \sum_{\mathcal{L}\subset\mathcal{K},\, \mathcal{B}_{\mathcal{L}}\neq\emptyset} \kappa(S_{\mathcal{L}}), \tag{34}$$

which means the function $S \mapsto \sum_{T\subseteq S,\, T\in\mathcal{B}_{\mathcal{K}}} \kappa(T)$ is constant on $\mathcal{B}_{\mathcal{K}}$. This allows one to derive (i) for $\mathcal{K}$, for instance, by induction on $|S|$ for $S \in \mathcal{B}_{\mathcal{K}}$. If we apply (34) to $S = S_{\mathcal{K}}$ we get (ii) for $\mathcal{K}$. $\square$

Example 14 Take $N = \{a, b, c\}$ and classify types of considered classes $\mathcal{A}$, specified by $\mathcal{A}_{\min}$. Using (30) we get the corresponding inequalities (28):

- $\mathcal{A}_{\min} = \{abc\}$ leads to $\kappa_{\mathcal{A}}(abc) = 1$ and $\kappa_{\mathcal{A}}(S) = 0$ otherwise. This gives the constraint $0 \leq \mathsf{c}(abc)$,

- $\mathcal{A}_{\min} = \{ab\}$ leads to $\kappa_{\mathcal{A}}(ab) = 1$ (and $\kappa_{\mathcal{A}}(S) = 0$ otherwise), which gives the constraint $0 \leq \mathsf{c}(ab)$,

- $\mathcal{A}_{\min} = \{ab, ac\}$ leads to $\kappa_{\mathcal{A}}(ab) = \kappa_{\mathcal{A}}(ac) = 1$ and $\kappa_{\mathcal{A}}(abc) = -1$, which gives the constraint $0 \leq \mathsf{c}(ab) + \mathsf{c}(ac) - \mathsf{c}(abc)$,

- $\mathcal{A}_{\min} = \{ab, ac, bc\}$ leads to $\kappa_{\mathcal{A}}(ab) = \kappa_{\mathcal{A}}(ac) = \kappa_{\mathcal{A}}(bc) = 1$ and $\kappa_{\mathcal{A}}(abc) = -2$, which gives the constraint $0 \leq \mathsf{c}(ab) + \mathsf{c}(ac) + \mathsf{c}(bc) - 2\mathsf{c}(abc)$,

- $\mathcal{A}_{\min} = \{c\}$ leads to $\kappa_{\mathcal{A}}(c) = 1$ which gives $0 \leq \mathsf{c}(c)$, which is a vacuous constraint because of $\mathsf{c}(c) = 1$ implied by (10),

- $\mathcal{A}_{\min} = \{c, ab\}$ leads to $\kappa_{\mathcal{A}}(c) = \kappa_{\mathcal{A}}(ab) = 1$ and $\kappa_{\mathcal{A}}(abc) = -1$, and then to $0 \leq \mathsf{c}(c) + \mathsf{c}(ab) - \mathsf{c}(abc)$, which leads after the substitution $\mathsf{c}(c) = 1$ to $0 \leq 1 + \mathsf{c}(ab) - \mathsf{c}(abc)$,

- $\mathcal{A}_{\min} = \{a, b\}$ leads to $\kappa_{\mathcal{A}}(a) = \kappa_{\mathcal{A}}(b) = 1$ and $\kappa_{\mathcal{A}}(ab) = -1$, and then, after substituing $\mathsf{c}(i) = 1$, to $0 \leq 2 - \mathsf{c}(ab)$,

- $\mathcal{A}_{\min} = \{a, b, c\}$ leads to $\kappa_{\mathcal{A}}(a) = \kappa_{\mathcal{A}}(b) = \kappa_{\mathcal{A}}(c) = 1$, $\kappa_{\mathcal{A}}(ab) = \kappa_{\mathcal{A}}(ac) = \kappa_{\mathcal{A}}(bc) = -1$ and $\kappa_{\mathcal{A}}(abc) = 1$, which gives, after the substitution $\mathsf{c}(i) = 1$, $0 \leq 3 - \mathsf{c}(ab) - \mathsf{c}(ac) - \mathsf{c}(bc) + \mathsf{c}(abc)$.

Thus, we see that the non-vacuous constraints are identical with the transformed elementary $\boldsymbol{\eta}$-constraints – see Example 9.

4.1.4 Remarks on the characteristic transformation

Let us consider the characteristic transformation given by (13) – see § 3.3.2. It can be viewed as a mapping $\boldsymbol{\eta} \mapsto \boldsymbol{B}\boldsymbol{\eta}$, where $\boldsymbol{B}$ is an $m \times n$ matrix, whose entries $\boldsymbol{b}[\,S,(i|B)\,]$ are specifed as follows: for $|S| \geq 1$, $i \in N$, $B \subseteq N \setminus \{i\}$,

$$\boldsymbol{b}[\,S,(i|B)\,] = \delta(\, i \in S \;\&\; S \setminus \{i\} \subseteq B\,) \equiv \delta(\, S \subseteq \{i\} \cup B\,) - \delta(\, S \subseteq B\,)\,. \tag{35}$$

There is a close relation to the matrix $\boldsymbol{A}$ introduced in (16). Indeed, there exists an invertible unimodular $m \times m$ matrix $\boldsymbol{C}$ such that $\boldsymbol{B} = \boldsymbol{C}\boldsymbol{A}$. More specifically, the entries $\boldsymbol{c}[\,S,T\,]$ of $\boldsymbol{C}$ for non-empty sets $S, T \subseteq N$ are given by

$$\boldsymbol{c}[\,S,T\,] = \begin{cases} \delta(\, S \subseteq T\,) & \text{if } |S| \geq 2, \\ \delta(\, S = T\,) & \text{if } |S| = 1. \end{cases}$$

To see it write for fixed $S \subseteq N$, $|S| \geq 2$ and a pair $(i|B)$ with help of (16):

$$\begin{aligned} \sum_{T \neq \emptyset} \boldsymbol{c}[\,S,T\,] \cdot \boldsymbol{a}[\,T,(i|B)\,] &= \sum_{T \supseteq S} \boldsymbol{a}[\,T,(i|B)\,] = \sum_{T \supseteq S} [\,\delta_{\{i\}\cup B}(T) - \delta_B(T)\,] \\ &= \sum_{T \supseteq S} \delta_{\{i\}\cup B}(T) - \sum_{T \supseteq S} \delta_B(T) \\ &= \delta(S \subseteq \{i\} \cup B) - \delta(S \subseteq B) = \boldsymbol{b}[\,S,(i|B)\,]\,. \end{aligned}$$

Analogously, for $S \subseteq N$, $|S| = 1$ one has

$$\sum_{T \neq \emptyset} \boldsymbol{c}[\,S,T\,] \cdot \boldsymbol{a}[\,T,(i|B)\,] = \sum_{T = S} \boldsymbol{a}[\,T,(i|B)\,] = \boldsymbol{a}[\,S,(i|B)\,] = \delta_{\{i\}}(S) = \boldsymbol{b}[\,S,(i|B)\,]\,.$$

We leave to the reader to verify that the $m \times m$-matrix $\boldsymbol{D}$ with entries $\boldsymbol{d}[\,T,R\,]$ for non-empty $T, R \subseteq N$ given by

$$\boldsymbol{d}[\,T,R\,] = \begin{cases} \delta(\, T \subseteq R\,) \cdot (-1)^{|R \setminus T|} & \text{if } |T| \geq 2, \\ \delta(\, T = R\,) & \text{if } |T| = 1. \end{cases}$$

is an inverse matrix to $\boldsymbol{C}$. Since both $\boldsymbol{C}$ and its inverse $\boldsymbol{D}$ are integral matrices, they are both unimodular. The following observation appears to be important.

Lemma 15 *Both the matrix $\boldsymbol{A}$ given by (16) and the matrix $\boldsymbol{B}$ given by (35) are full row rank unimodular matrices.*

Proof. Since $\boldsymbol{A} = \boldsymbol{D}\boldsymbol{B}$ where $\boldsymbol{D}$ is an invertible unimodular $m \times m$-matrix, it is enough to show that $\boldsymbol{B}$ is unimodular. By Proposition 12 and $\boldsymbol{B} = \boldsymbol{C}\boldsymbol{A}$ we already know that $\boldsymbol{B}$ has full row rank.

To show it is unimodular we re-label its columns and add some new ones. The original columns of $\boldsymbol{B}$ corresponding to pairs $(i|B)$ with $B \neq \emptyset$ are re-labelled by pairs $(C:B)$ of sets $\emptyset \neq B \subseteq C \subseteq N$ with $|C \setminus B| = 1$; that is, $(i|B)$ is replaced by $(C:B)$ where $C = \{i\} \cup B$. The formula (35) implies

$$\boldsymbol{b}[\,S,(C:B)\,] = \delta(\, S \subseteq C\,) - \delta(\, S \subseteq B\,) \qquad \text{for } S \subseteq N,\ |S| \geq 1.$$

The original column corresponding to a pair $(i|\emptyset)$, $i \in N$ is re-labelled by a singleton set $R = \{i\}$. Note that the column has the form δ_R. The newly added columns are labelled by sets $R \subseteq N$, $|R| \geq 2$ and defined as follows:

$$\boldsymbol{b}\,[\,S, R\,] = \delta(\,S \subseteq R\,) \qquad \text{for } S \subseteq N,\ |S| \geq 1.$$

Observe that this formula also holds in case $|R| = 1$. Now, it is enough to show that the extended matrix $\boldsymbol{B}$ is unimodular.

Let $\bar{\boldsymbol{B}}$ denote the $m \times m$-submatrix of $\boldsymbol{B}$ corresponding to columns labelled by sets $\emptyset \neq R \subseteq N$. It follows from the above description of columns in $\boldsymbol{B}$ that $\boldsymbol{B} = \bar{\boldsymbol{B}}\boldsymbol{E}$ where the matrix $\boldsymbol{E}$ has the entries $\boldsymbol{e}\,[\,T, R\,]$ for $\emptyset \neq T, R \subseteq N$ and $\boldsymbol{e}\,[\,T, (C:B)\,]$ for $\emptyset \neq T \subseteq N$, $\emptyset \neq B \subseteq C \subseteq N$, $|C \setminus B| = 1$ specified as follows:

$$\begin{array}{lcl} \boldsymbol{e}\,[\,T, R\,] & = & \delta(\,T = R\,)\,, \\ \boldsymbol{e}\,[\,T, (C:B)\,] & = & \delta(\,T = C\,) - \delta(\,T = B\,)\,. \end{array}$$

Therefore, it is enough to show that $\bar{\boldsymbol{B}}$ is invertible unimodular matrix and $\boldsymbol{E}$ totally unimodular (cf. Theorem 21.6 in [9]). We leave to the reader to verify that the inverse matrix $\boldsymbol{F}$ to $\bar{\boldsymbol{B}}$ has the entries

$$\boldsymbol{f}\,[\,R, U\,] = \delta(\,R \subseteq U\,) \cdot (-1)^{|U \setminus R|} \qquad \text{for } \emptyset \neq R, U \subseteq N.$$

Since $\bar{\boldsymbol{B}}$ has integral inverse $\boldsymbol{F}$, it is unimodular. The matrix $\boldsymbol{E}$ is totally unimodular because it is the restriction of a network matrix (cf. § 19.3 of [9]). More specifically, one can add one dummy row to $\boldsymbol{E}$, labelled by $S = \emptyset$: put $\boldsymbol{e}\,[\,\emptyset, R\,] = -1$ for $\emptyset \neq R \subseteq N$ and $\boldsymbol{e}\,[\,\emptyset, (C:B)\,] = 0$ for any pair $(C:B)$. We obtain a matrix with entries in $\{-1, 0, +1\}$ such that each of its columns contains exactly once $+1$ and exactly once -1. As mentioned in the statement (18) of § 19.3 in [9], such a matrix is totally unimodular. Of course, it remains totally unimodular if the row corresponding to $S = \emptyset$ is again removed. $\square$

4.2 Transformation of cluster inequalities

Luckily, these inequalites transform nicely to the framework of imsets.

Lemma 16 *Provided $\boldsymbol{\eta}$ satisfies (2), the cluster inequality (3) for $C \subseteq N$, $|C| \geq 2$ can be re-written either in terms of* u*-vectors as*

$$\sum_{T \subseteq N,\, |C \cap T| \geq 2} \mathsf{u}(T) \cdot (|C \cap T| - 1) \geq 0\,, \tag{36}$$

or in terms of c*-vectors as*

$$|C| - 1 - \sum_{S \subseteq C,\, |S| \geq 2} \mathsf{c}(S) \cdot (-1)^{|S|} \geq 0\,. \tag{37}$$

Proof. By (3), it is enough to show that the following equalities hold

$$\underbrace{|C| - \sum_{S\subseteq C,\,|S|\geq 2} \mathsf{c}(S)\cdot(-1)^{|S|}}_{\equiv(*)} \;=\; 1 + \sum_{T\subseteq N,\,|C\cap T|\geq 2} \mathsf{u}(T)\cdot(|C\cap T|-1)$$

$$= \sum_{i\in C} \; \sum_{B\subseteq N\setminus\{i\},\,B\cap C=\emptyset} \eta(i|B)\,.$$

Let $(*)$ denote the first expression there and write by (9)-(8):

$$(*) = |C| - \sum_{S\subseteq C,\,|S|\geq 2} (-1)^{|S|}\cdot[\,1 - \sum_{T\supseteq S}\mathsf{u}(T)\,]$$

$$= \;|C| - \underbrace{\sum_{S\subseteq C,\,|S|\geq 2} (-1)^{|S|}}_{|C|-1} + \sum_{S\subseteq C,\,|S|\geq 2} (-1)^{|S|}\cdot\sum_{T\supseteq S}\mathsf{u}(T)$$

$$= \;1 + \sum_{S\subseteq C,\,|S|\geq 2}\;\sum_{T\supseteq S}\mathsf{u}(T)\cdot\;(-1)^{|S|} = 1 + \sum_{T,\,|C\cap T|\geq 2}\mathsf{u}(T)\cdot\underbrace{\sum_{S\subseteq C\cap T,\,|S|\geq 2}(-1)^{|S|}}_{|C\cap T|-1}\,.$$

This already proves the first equality. Now, we substitute (7) in the last expression (note $|T|\geq 2$ for T here) and change the order of summation:

$$(*) \;=\; 1 + \sum_{T,\,|C\cap T|\geq 2}\mathsf{u}(T)\cdot(|C\cap T|-1)$$

$$= \;1 + \overbrace{\sum_{T,\,|C\cap T|\geq 2}\delta_N(T)\cdot(|C\cap T|-1)}^{|C|-1} + \sum_{i\in N}\;\sum_{B\subseteq N\setminus\{i\}}\eta(i|B)\cdot$$

$$\{\sum_{T,\,|C\cap T|\geq 2}\delta_B(T)\cdot(|C\cap T|-1) - \sum_{T,\,|C\cap T|\geq 2}\delta_{\{i\}\cup B}(T)\cdot(|C\cap T|-1)\,\}$$

$$= \;|C| + \sum_{i\in N}\;\sum_{B\subseteq N\setminus\{i\}}\eta(i|B)\cdot$$

$$\{\,\delta(\,|C\cap B|\geq 2\,)\cdot(|C\cap B|-1) - \delta(\,|C\cap(\{i\}\cup B)|\geq 2\,)\cdot(|C\cap(\{i\}\cup B)|-1)\,\}\,.$$

Now, we realize the that the inner expression in braces vanishes for $i\notin C$ because then $C\cap B = C\cap(\{i\}\cup B)$. Analogously, it vanishes if $i\in C$ but $C\cap B=\emptyset$. However, in case $i\in C$ and $C\cap B\neq\emptyset$ it equals to -1. Thus, we write using (2) for $i\in C$:

$$(*) \;=\; |C| + \sum_{i\in N}\;\sum_{B\subseteq N\setminus\{i\}}\eta(i|B)\cdot\delta(\,i\in C\;\&\;C\cap B\neq\emptyset\,)\cdot(-1)$$

$$= \;|C| - \sum_{i\in C}\;\sum_{B\subseteq N\setminus\{i\}}\eta(i|B)\cdot\delta(\,C\cap B\neq\emptyset\,)$$

$$\stackrel{(2)}{=} \;\sum_{i\in C}\;\sum_{B\subseteq N\setminus\{i\}}\eta(i|B) - \sum_{i\in C}\;\sum_{B\subseteq N\setminus\{i\},\,B\cap C\neq\emptyset}\eta(i|B)$$

$$= \;\sum_{i\in C}\;\sum_{B\subseteq N\setminus\{i\},\,B\cap C=\emptyset}\eta(i|B)\,,$$

which gives the third required equality. $\square$

Example 17 Take $N = \{a, b, c\}$. By (36), there are four transformed cluster inequalities for $|C| \geq 2$ breaking into two types:

- $\mathsf{u}(\{a,b\}) + \mathsf{u}(\{a,b,c\}) \geq 0,$ (for $C = \{a,b\}$)
- $\mathsf{u}(\{a,b\}) + \mathsf{u}(\{a,c\}) + \mathsf{u}(\{b,c\}) + 2 \cdot \mathsf{u}(\{a,b,c\}) \geq 0.$ (for $C = \{a,b,c\}$)

We observe they coincide with two types of non-specific inequality constraints mentioned in Example 5 (see § 3.2.1). Nevertheless, the remaining non-specific constrain mentioned there, namely $\mathsf{u}(\{a,b,c\}) \geq 0$, is not implied by the transformed cluster inequalities. For instance, the u-vector given by $\mathsf{u}(T) = (-1)^{|T|}$ for $T \subseteq \{a,b,c\}$ shows that.

The above example suggests that the transformed cluster inequalities are implied by the non-specific ones, which is indeed the case.

Corollary 18 *The cluster inequalities transformed to the framework of u-vectors (36) follow from non-specific inequality constraints (6).*

Proof. By (36), the cluster inequality for $C \subseteq N$, $|C| \geq 2$ has the form

$$\langle m_C, \mathsf{u} \rangle = \sum_{T \subseteq N} m_C(T) \cdot \mathsf{u}(T) \geq 0, \quad \text{with } m_C(T) = \max\{0, |C \cap T| - 1\} \text{ for } T \subseteq N.$$

The function m_C is a special (standardized extreme) supermodular function, and, therefore, the inequality for C follows from (6). $\square$

Thus, we can summarize. The exact translation of the equality constraints (2) to the framework of u-vectors are the equality constraints (4) – see § 3.2.2. Provided (2) is valid, the exact translation of non-negativity constraints (1) are specific inequality constraints (5) (see Corollary 11 in § 4.1.1), and by Corollary 18, the cluster inequalities (3) translate to some of the non-specific inequality constraints (6). In particular, we have

Corollary 19 *The u-polyhedron specified by (4)-(6) is contained in the image of the $\boldsymbol{\eta}$-polyhedron specified by (1)-(3) by the mapping $\boldsymbol{\eta} \mapsto \mathsf{u}^{\boldsymbol{\eta}}$ defined in (7), which is the polyhedron specified by (4), (5) and (36).*

5 LP relaxation

To motivate the next result consider the case of three variables and transform Jaakkola et al.'s polyhedron J (§ 3.1.1) to the framework of c-vectors.

Example 20 In Example 9 (see § 4.1), we transformed the elementary constraints (1)-(2) to the framework of characteristic imsets in case $N = \{a, b, c\}$. The result was a polyhedron given by fifteen inequalities and four equality constraints. One can add the transformed cluster inequalities (37) to those constraints. There are four such inequalities breaking into two types:

- $0 \leq 1 - \mathsf{c}(ab)$, (for $C = \{a, b\}$)
- $0 \leq 2 - \mathsf{c}(ab) - \mathsf{c}(ac) - \mathsf{c}(bc) + \mathsf{c}(abc)$. (for $C = \{a, b, c\}$)

We computed (again by *Polymake* [3]) the vertices of the resulting polyhedron (= the image of J) and got 12 vertices. The type representatives are as follows:

$$[0, 0, 0, 0], [1, 0, 0, 0], [1, 1, 0, 0], [1, 1, 0, 1], [1, 1, 1, 1], [1, 1, 1, \frac{3}{2}].$$

All the eleven lattice points here are characteristic imsets (for acyclic directed graphs), while the fractional vertex $[1, 1, 1, \frac{3}{2}]$ is not. However, it is a convex combination of vertices of the bigger polyhedron (= of the image of J'), namely of $[2, 2, 2, 3]$ and $[0, 0, 0, 0]$ – see Example 9.

To get the exact polyhedral characterization of the characteristic imset polytope (= of the convex hull of the set of characteristic imsets) in this case $N = \{a, b, c\}$ one has to add the translation of the non-specific inequality constrain $\mathsf{u}(\{a, b, c\}) \geq 0$ – see Example 17. By (8)-(9), it leads to

- $\mathsf{c}(abc) \leq 1$,

which clearly cuts off the fractional vertex and the result is just the polytope spanned by the remaining eleven lattice points. Thus, the example shows that the basic inequalities for characteristic imsets mentioned in § 3.3.1 (Corollary 6) are not implied by the transformed Jaakkola et al.'s inequalities (1)-(3).

Nevertheless, we have observed that in case $|N| = 3$ the only lattice points within the transformed polyhedron are the characteristic imsets. This leads to a hypothesis that this holds for any $|N|$. We confirm this conjecture now using the observation from Lemma 15. Thus, by transforming Jaakkola et al.'s polyhedron J we get an explicit LP relaxation of the characteristic imset polytope.

Corollary 21 *The only lattice points within the polyhedron of* c*-vectors given by (10), (28) and (37) are characteristic imsets (for acyclic directed graphs).*

Proof. Let us interpret any c-vector as an element of $\mathbb{R}^{\mathcal{P}_1(N)} \equiv \mathbb{R}^m$, that is, $\mathsf{c}(\emptyset) = 1$ by a convention. We have already observed that the polyhedron given by (10), (28) and (37) is the image of the polyhedron J specified by (1)-(3) by the transformation $\boldsymbol{\eta} \mapsto \boldsymbol{B}\boldsymbol{\eta} = \mathsf{c}$ defined in (13) – see § 4.1.3 and § 4.2.

Assume c is a lattice point in the considered polyhedron. Thus, c has a preimage $\boldsymbol{x} \in \mathsf{J}$, which implies that the polyhedron $\{\boldsymbol{x} \geq 0;\ \boldsymbol{B}\boldsymbol{x} = \mathsf{c}\}$ in $\mathbb{R}^n$ is non-empty. By Lemma 15, $\boldsymbol{B}$ is unimodular, which allows us to use Theorem 19.2 in [9] saying that a full row rank $m \times n$-matrix $\boldsymbol{B}$ is unimodular if and only if the polyhedron $\{\boldsymbol{x} \geq 0;\ \boldsymbol{B}\boldsymbol{x} = \mathsf{c}\}$ is integral for any $\mathsf{c} \in \mathbb{Z}^m$. That means, it is the convex hull of its lattice points. In particular, since it is non-empty, it has at least one lattice point. Let us fix one such lattice point $\boldsymbol{\eta} \in \mathbb{Z}^n$, $\boldsymbol{\eta} \geq 0$ with $\boldsymbol{B}\boldsymbol{\eta} = \mathsf{c}$. It automatically satisfies (1); (2) holds because $\mathsf{c}(S) = 1$ for $|S| = 1$ and $\boldsymbol{B}\boldsymbol{\eta} = \mathsf{c}$. As (37) holds for c, $\boldsymbol{\eta}$ satisfies all cluster inequalities (3) (by Lemma 16). That means, $\boldsymbol{\eta}$ is a lattice point in J.

By Lemma 4, $\boldsymbol{\eta}$ is necessarily the code $\boldsymbol{\eta}_G$ of an acyclic directed graph G over N. By Lemma 7, its image c by the characteristic transformation is the characteristic imset c_G corresponding to G. □

Remark In the proof of Corollary 21, we have shown that if c is a lattice point in the cone generated by columns of $\boldsymbol{B}$ then it is a non-negative integer combination of columns of $\boldsymbol{B}$. That means, in terms of § 16.4 of [9], the columns of $\boldsymbol{B}$ form the minimal integral Hilbert basis of the cone generated by them. Following the terminology from commutative algebra, the semigroup generated by columns of $\boldsymbol{B}$ is *normal* [6, 15, 16].

Nevertheless, because of the one-to-one correspondence between u-vectors and c-vectors, we have an analogous result in the framework of standard imsets.

Corollary 22 *The polyhedron of* u*-vectors given by (4), (5) and (36) is an LP relaxation of the standard imset polytope.*

Proof. As explained in § 3.3 the mapping $\mathsf{u} \mapsto \mathsf{c}$ given by (8)-(9) is invertible and maps lattice points to lattice points. Moreover, (4) is transformed to (10), (5) to (28) by Lemma 13 and (36) to (37) by Lemma 16. Thus, the image of the polyhedron is the polyhedron of c-vectors from Corollary 21. The pre-images of characteristic imsets are standard imsets. □

Note that one can also prove Corollary 22 directly, by the method Corollary 21 was proved. Indeed, one can use an analogous consideration where the matrix $\boldsymbol{B}$ is replaced by $\boldsymbol{A}$ and the vector c by $\boldsymbol{b}_{\mathsf{u}}$ for an u-vector – see the relation (17) mentioned in § 4.1.1.

Thus, we have an explicit LP relaxation of the standard imset polytope and the conjecture from [14] is confirmed:

Corollary 23 *The polyhedron of* u*-vectors given by (4)-(6) is an LP relaxation of the standard imset polytope.*

Proof. This follows from Corollaries 22 and 19. □

Conclusions

Corollary 21 gives an explicit LP relaxation of the characteristic imset polytope. Nevertheless, some of the inequalities (28) are superfluous because they follow from the remaining inequalities. Moreover, perhaps adding the basic inequalities from Corollary 6 allows one further reduction of the number of inequalities.

Another research direction is to look for even more loose LP relaxation of the standard/characteristic imset polytope, which however, has a less number of inequalities.

Acknowledgments

Milan Studený was supported by GA ČR grant 201/08/0539.

References

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

[2] D. M. Chickering: Optimal structure identification with greedy search, Journal of Machine Learning Research 23 (2002) 507-554.

[3] E. Gawrilow, M. Joswig: *Polymake* – a framework for analyzing convex polytopes, Polytopes – Combinatorics and Computation (Oberwolfach, 1997), DMV Seminar 29, Birkhäuser 2000.

[4] T. Jaakkola, D. Sontag, A. Globerson, M. Meila: Lerning Bayesian network structure using LP relaxations, in Proceedings of the 13th International Conference on Artificial Intelligence and Statistics 2010, Chia Laguna Resort, Sardinia, Italy, pp. 358-365.

[5] S. L. Lauritzen: Graphical Models, Clarendon Press 1996.

[6] E. Miller, B. Sturmfels: Combinatorial Commutative Algebra, Springer Verlag 2005.

[7] R. E. Neapolitan: Learning Bayesian Networks, Pearson Prentice Hall 2004.

[8] J. Pearl: Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann 1988.

[9] A. Schrijver: Theory of Linear and Integer Programming, John Wiley 1986.

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

[11] M. Studený: Probabilistic Conditional Independence Structures, Springer Verlag 2005.

[12] M. Studený, J. Vomlel, R. Hemmecke: A geometric view on learning Bayesian network structures, International Journal of Approximate Reasoning 51 (2010) 578-586.

[13] M. Studený, R. Hemmecke, S. Lindner: Characteristic imset: a simple algebraic representative of a Bayesian network structure, in Proceedings of the 5th European Workshop on Probabilistic Graphical Models (P. Myllymäki, T. Roos, T. Jaakkola eds.), HIIT Publications 2010, 257-264.

[14] M. Studený, J. Vomlel: On open questions in the geometric approach to structural learning Bayesian nets, International Journal of Approximate Reasoning 52 (2011) 627-640.

[15] B. Sturmfels: Gröbner Bases and Convex Polytopes, American Mathematical Society 1995.

[16] A. Takemura, R. Yoshida: A generalization of the integer linear infeasibility problem, Discrete Optimization 5 (2007) 36-52.

[17] 4ti2 team: *4ti2* – a software package for algebraic, geometric and combinatorial problems on linear spaces. Available eletronically at `www.4ti2.de`, 2008.