Towards using the chordal graph polytope in learning decomposable models

Milan Studený a,*, James Cussens b

a Institute of Information Theory and Automation of the CAS, Prague, Pod Vodárnu věži 4, 18208, Czech Republic
b Dept of Computer Science & York Centre for Complex Systems Analysis, University of York, Deramore Lane, York, YO10 5GH, United Kingdom

Abstract

The motivation for this paper is the integer linear programming approach to learning the structure of a decomposable graphical model. We have chosen to represent decomposable models by means of special zero-one vectors, named characteristic imsets. Our approach leads to the study of a special polytope, defined as the convex hull of all characteristic imsets for chordal graphs, named the chordal graph polytope. In this theoretical paper, we introduce a class of clutter inequalities (valid for the vectors in the polytope) and show that all of them are facet-defining for the polytope. We dare to conjecture that they lead to a complete polyhedral description of the polytope. Finally, we propose a linear programming method to solve the separation problem with these inequalities for the use in a cutting plane approach.

Keywords: learning decomposable models, integer linear programming, characteristic imset, chordal graph polytope, clutter inequalities, separation problem

1. Introduction: explaining the motivation

Decomposable models are fundamental probabilistic graphical models [16]. A well-known fact is that elegant mathematical properties of these structural models form the theoretical basis of the famous method of local computation [6]. Decomposable models, which are described by chordal undirected graphs, can be viewed as special cases of Bayesian network models [19], which are described by directed acyclic graphs.

Two traditionally separate disciplines in probabilistic graphical models are learning and inference. Structure learning is determining the graphical model, represented by a graph, on the basis of observed statistical data. Inference in Bayesian network models has two phases. The first one is transformation of the (learned) directed acyclic graph into a junction tree, which can be viewed as a representative of a decomposable model. The second phase in inference is proper local computation (of conditional probabilities) in a junction tree. The motivation for the present paper is the idea to merge structural learning with the junction tree construction in one step, which basically means direct learning the structure of a decomposable model on basis of data.

There are various methods for learning decomposable model structure, most of them being specializations of the methods for learning Bayesian network structure [18]. There are methods based on statistical conditional independence tests like the PC algorithm [23] or MCMC simulations [11]. This particular paper deals with a score-based approach, where the task is to maximize some additively decomposable score, like the BIC score [21] or the BDeu score [12]. There are some arguments in favour of this approach in comparison with the methods based on statistical tests. Specifically, the study in [29] indicates that some classes of domain models cannot be learned by procedures that modify the graph structure by one edge at a time.

We are interested in the integer linear programming (ILP) approach to structural learning of decomposable models. The idea behind this approach is to encode graphical models by certain vectors with integer components in such a way that the usual scores become linear or affine functions of the vector representatives. There are several ways to encode Bayesian network models; the most successful one seems to be to encode them by family-variable vectors as used in [14], [17] and [11]. On the other hand, since the present paper deals with learning decomposable models...
we have intentionally chosen to encode them by different vector representatives, called characteristic imsets; these
vectors have also been applied in the context of learning Bayesian network models in [13] and [26]. This mode of
representation leads to an elegant and unique way of encoding decomposable models which we believe is particularly
suitable for structure learning of these models.

Let us note that two recent conference papers have also been devoted to ILP-based learning of decomposable
models, but they use different binary encodings of the models. More specifically, Sesh Kumar and Bach [22] used
special codes for junction trees of the graphs, while Pérez et al. [20] encoded certain special coarsenings of maximal
hyper-trees. Moreover, the goal in both these papers was learning a specifically restricted class of decomposable
models (namely, all cliques have the same prescribed size and the same holds for separators) unlike in this theoretical
paper, which we hope to be the first step towards a general ILP method for learning decomposable models.

Two other recent papers devoted to structure learning of decomposable models also used encodings of junction
trees. Corander et al. [5] expressed the search space in terms of logical constraints and used constraint satisfaction
solvers. Even better running times have been achieved by Kangas et al. [15], who applied the idea of decomposing
junction trees into subtrees, which allowed them to use the method of dynamic programming. We note that the
junction tree representation is closely related to the (superset) Möbius inversion of the characteristic imset we mention
in Section 6.

Our approach leads to the study of the geometry of a polytope defined as the convex hull of all characteristic imsets
for chordal graphs (over a fixed set of nodes \(N\)), with the possible modification that a clique size limit is given. This
polytope has already been dealt with by Lindner [17] in her thesis, where she derived some basic observations on the
polytope. For example, she mentioned that a complete facet description of the polytope with cliques size limit two,
which corresponds to learning undirected forests, can be derived. She also identified some non-trivial inequalities
for the polytope with no clique size limit. Being inspired by Lindner we name this polytope the “chordal graph
characteristic imset polytope”, but abbreviate this to the chordal graph polytope.

In this paper, which is an extended version of a proceedings paper [23], we assume that the reader is familiar
with basic concepts of polyhedral geometry, as presented in numerous textbooks on this topic; for example in [2] or
[28]. We present a complete facet description of the polytope where \(|N| \leq 4\) and mention the case \(|N| = 5\), where the
facet description is also available. We have succeeded in classifying all facet-defining inequalities for this polytope in
these cases. What we found out is that, with the exception of a natural lower bound inequality, there is a one-to-one
correspondence between the facet-defining inequalities and the clutters (alternatively named antichains or Sperner
families) of subsets of the variable set \(N\) (i.e. of the set of nodes) containing at least one singleton; so we call these
clutter inequalities.

This establishes a sensible conjecture about a complete polyhedral description of the polytope (with no clique
size limit). We prove that every clutter inequality is both valid and facet-defining for the polytope. We also tackle an
important separation problem: that is, given a non-integer solution to a linear programming (LP) relaxation problem,
find a clutter inequality which (most) violates the current solution.

The structure of the paper

Basic concepts are recalled in Section 2 where the concept of a chordal graph polytope is introduced. Section 3
reports on the facet description of this polytope in the case of at most five nodes, which was found computationally.
The clutter inequalities are defined and illustrated by a couple of simple examples in Section 4. Our completeness
conjecture is then formulated in Section 5; various other versions of clutter inequalities are given in Section 6.
In Section 7 we present the idea of the proof of their validity for any vector in the chordal graph polytope. The main result
of the paper saying that every clutter inequality is facet-defining for the polytope is presented in Section 8. Section 9
is devoted to a sub-problem of finding a suitable clutter inequality in the context of the cutting plane method. A brief
report on a small preliminary empirical study is given in Section 10. Important open tasks are recalled in Conclusions,
which is Section 11. The proofs of most observations have been put in the Appendix to make the paper smoothly
readable.

2. Basic concepts

Let \(N\) be a finite set of variables; assume that \(n := |N| \geq 2\) to avoid the trivial case. In the statistical context, the
elements of \(N\) correspond to random variables, while in the graphical context they correspond to nodes of graphs.
2.1. Some conventional notation and terminology

The symbol \( \subset \) will be used to denote non-strict set inclusion of unlike \( \subseteq \), which will serve to denote strict inclusion: \( S \subset T \) means \( S \subseteq T \) and \( S \neq T \). The power set of \( N \) will be denoted by \( \mathcal{P}(N) := \{ S : S \subseteq N \} \).

We are going to use the term clutter to name any collection \( \mathcal{L} \) of subsets of \( N \) that are inclusion incomparable, that is, \( L, R \in \mathcal{L} \) and \( L \subseteq R \) implies \( L = R \). Such set collections are alternatively named Sperner families or antichains in the mathematical literature. Occasionally, we will abbreviate notation for the union of sets in a clutter: \( \bigcup \mathcal{L} := \bigcup_{L \in \mathcal{L}} L \).

Given a clutter \( \mathcal{L} \) of subsets of \( N \) such that \( \emptyset \neq \bigcup \mathcal{L} \), we introduce the notation \( \mathcal{L}^{\uparrow} \) for the filter generated by \( \mathcal{L} \), by which is meant a system of subsets of \( N \) closed under supersets:

\[
\mathcal{L}^{\uparrow} := \{ T \subseteq N : \exists L \in \mathcal{L} \quad L \subseteq T \}.
\]

Moreover, we are going to use special notation for the zero-one indicator of a predicate \( P \):

\[
\delta(P) := \begin{cases} 
1 & \text{if the predicate } P \text{ holds,} \\
0 & \text{if } P \text{ does not hold.}
\end{cases}
\]

The abbreviation LHS will mean “left-hand side” (of an inequality), while RHS will be a shorthand for “right-hand side”. The symbol \( T := \{ S \subseteq N : |S| \geq 2 \} \) will be our notation for the collection of non-empty non-singletons, used as a standard index set for components of our vectors.

2.2. Chordal undirected graphs

We say that a graph \( G \) is over \( N \) if \( G \) has \( N \) as the set of nodes and it is undirected if every edge is undirected. An undirected graph is chordal if every cycle of length at least 4 has a chord, that is, an edge connecting non-consecutive nodes in the cycle. A set \( S \subseteq N \) is complete if every two distinct nodes in \( S \) are connected by an edge. Maximal complete sets with respect to inclusion are called the cliques (of \( G \)). A well-known equivalent definition of a chordal graph is that the collection of its cliques can be ordered into a sequence \( C_1, \ldots, C_m \), \( m \geq 1 \), satisfying the running intersection property (RIP):

\[
\forall i \geq 2 \exists j < i \text{ such that } S_i := C_i \cap \left( \bigcup_{j \neq i} C_j \right) \subseteq C_j.
\]

The sets \( S_i = C_i \cap (\bigcup_{j \neq i} C_j) \), \( i = 2, \ldots, m \), are the respective separators. The multiplicity of a separator \( S \) is the number of indices \( 2 \leq i \leq m \) such that \( S = S_i \); the separators and their multiplicities are known not to depend on the choice of the ordering satisfying the RIP, see [24, Lemma 7.2]. Each chordal graph defines the respective statistical decomposable model; see [10, § 4.4].

2.3. Learning graphical models

The score-based approach to structure learning of graphical models is based on maximizing some scoring criterion, briefly called a score, which is a bivariate real function \( (G, D) \mapsto Q(G, D) \) of the graph \( G \) and the (observed) database \( D \). In the context of learning Bayesian networks, that is, graphical models described by directed acyclic graphs \( H \), a crucial technical assumption [3] is that \( Q \) should be additively decomposable, which means, it has the form

\[
Q(H, D) = \sum_{a \in N} q_D(a | \text{pa}_H(a))
\]

where the summands \( q_D(\cdot | \cdot) \) are called local scores, and \( \text{pa}_H(a) := \{ b \in N : b \to a \text{ in } H \} \) is the set of parents of a node \( a \) in \( H \). All criteria used in practice satisfy this requirement, as long as the data contain no missing values. Another typical assumption is that \( Q \) is score equivalent [3], which means that Markov equivalent (directed acyclic) graphs yield the same score. In the present paper, we are going to adopt this approach to learning chordal undirected graphical models.
2.4. Characteristic imset

The concept of a characteristic imset (for a directed acyclic graph) was introduced in [13]. Each characteristic imset is an element of the real vector space $\mathbb{R}^\mathcal{V}$ where $\mathcal{V} = \{S \subseteq N : |S| \geq 2\}$ is the collection of non-empty non-singletons. A fundamental fact is that every additively decomposable and score equivalent scoring criterion turns out to be an affine function (that is, a linear function plus a constant) of the characteristic imset encoding the graph. These special zero-one vectors describe uniquely the equivalence classes of directed acyclic graphs.

Nevertheless, in the sub-frame of decomposable models, that is, in the frame of graphical models described by chordal undirected graphs, models are in a one-to-one correspondence with (chordal undirected) graphs and the next simpler definition can be used; see [13, Corollary 4].

Definition 1 (characteristic imset).

Given a chordal undirected graph $G$ over $N$, the characteristic imset of $G$ is a zero-one vector $c_G$ with components indexed by subsets $S$ of $N$ with $|S| \geq 1$:

$$c_G(S) = \begin{cases} 1 & \text{if } S \text{ is a complete set in } G, |S| \geq 1, \\ 0 & \text{for remaining } S \subseteq N, |S| \geq 1. \end{cases}$$

An implicit consequence is that $c_G(L) = 1$ for any graph $G$ over $N$ and any singleton $L \subseteq N, |L| = 1$.

Thus, the characteristic imset $c_G$ is basically an element of $\mathbb{R}^\mathcal{V}$ because its values for singletons are fixed to be 1. A conventional value $c_G(\emptyset)$ for the empty set can also be fixed, but it plays no substantial role in the theory because it does not occur in basic versions of our inequalities from Section 4. Nonetheless, we accept the convention $c_G(\emptyset) := 1$ in this paper because it leads to an elegant Möbius inversion formula (see Lemma 3 in Section 6.1). In particular, $c_G$ can be viewed as a vector in $\mathbb{R}^{P(N)}$.

As decomposable models induced by chordal undirected graphs can be viewed as special cases of Bayesian network models each sensible scoring criterion is an affine function of the characteristic imset. Specifically, [26, Lemma 3] implies that an additively decomposable and score equivalent criterion $Q$ has the form

$$Q(G, D) = k + \sum_{S \in \mathcal{V}} r^D_G(S) \cdot c_G(S),$$

where $k$ is a constant and,

for any $S \in \mathcal{V}$, $r^D_G(S) = \sum_{K \subseteq S, a \in S} (-1)^{|S\setminus K| + 1} \cdot q_D(a|K)$, with arbitrary $a \in S$.

where $q_D(*)$ are the respective local scores.

2.5. Chordal graph polytope

Now, we introduce the polytope to be studied. To be flexible, we consider the situation where a maximal clique size limit $k$ is given, $2 \leq k \leq n = |N|$. Taking $k = n$ gives the general (unrestricted) case while taking $k = 2$ leads to a well-known special case of undirected forests.

Definition 2 (chordal graph polytope).

The chordal graph polytope over $N$ with clique size at most $k$, where $2 \leq k \leq n = |N|$, is as follows:

$$D^k_N := \text{conv}\{(c_G : G \text{ chordal graph over } N \text{ with clique size at most } k)\},$$

where $\text{conv}(-)$ denotes the convex hull.

The dimension of the polytope $D^k_N$ is $\sum_{\ell=2}^k \binom{n}{\ell}$. Thus, for the unrestricted polytope $D_N := D^k_N$, one has dim ($D_N$) = $\sum_{\ell=2}^n \binom{n}{\ell} = 2^n - n - 1$, while one has dim ($D^k_N$) = $\binom{n}{k}$ for the polytope encoding undirected forests.

In fact, one can decide to be even more general and consider the polytope

$$D^K_N := \text{conv}\{(c_G : G \text{ chordal graph over } N \text{ with complete sets in } K)\}$$

for a general confining system $K \subseteq \mathcal{P}(N)$ of subsets of $N$ closed under subsets and containing (all) singletons in $N$. Our long-term strategic goal and the topic of future research is to get the facet description of $D^K_N$ for any such
confining system \( \mathcal{K} \) and utilize such result in learning decomposable models by ILP methods. The idea is that \( \mathcal{K} \) will be obtained as the result of a pruning procedure to be developed, which ensures that every optimal chordal graph has complete sets in \( \mathcal{K} \). The point is that dim \( (\mathcal{D}_N^\mathcal{K}) \) = \( |\mathcal{K}| - n - 1 \) can be considerably smaller than dim \( (\mathcal{D}_N) \). Let us note that the assumption on \( \mathcal{K} \) being closed under subsets is not restrictive because, given a general system \( \mathcal{K} \) containing sets \( L \subseteq N, |L| \leq 1 \), one has \( \mathcal{D}_N^\mathcal{K} = \mathcal{D}_N^\mathcal{K}' \) with \( \mathcal{K}' = \{ S \subseteq N : T \in \mathcal{K} \text{ for any } T \subseteq S \} \); this follows from the fact that, for any \( c \in \mathcal{D}_N \) and \( \emptyset \neq T \subseteq S \subseteq N \), the equality \( \varepsilon(T) = 0 \) implies \( \varepsilon(S) = 0 \).

3. Example: the cases of a low number of variables

For small values of \( n = |N| \) we have been able to use the cdd program \([10]\) to compute a facet description of the chordal graph polytope \( \mathcal{D}_N \). In the case \( n = 3 \), \( \mathcal{D}_N \) has 8 vertices, encoding 8 chordal graphs, and 8 facet-defining inequalities, decomposing into 4 permutation types.

- lower bound: \( 0 \leq \varepsilon((a, b, c)) \) (1 inequality),
- 2-to-3 monotonicity inequalities: \( \varepsilon((a, b, c)) \leq \varepsilon((a, b)) \) (3 inequalities),
- upper bounds: \( \varepsilon((a, b)) \leq 1 \) (3 inequalities),

cluster inequality for a 3-element set: \( \varepsilon((a, b)) \leq \varepsilon((a, c)) + \varepsilon((c, b)) + \varepsilon((a, c)) + \varepsilon((b, c)) \leq 2 + \varepsilon((a, b, c)) \) (1 inequality).

Note that the cluster inequalities (formulated in terms of family variables) have earlier occurred in the context of learning Bayesian networks; see Example 3 in Section 6.2 and references \([14], [1], [26]\). The restricted polytope \( \mathcal{D}_N^3 \) has this form: here \( \gamma = c \), \( \kappa((a, b)) = 1 \) and \( \kappa(S) = 0 \) for \( S \subset [a, b] \). The 3-element cluster inequality for \( n = 3 \) can be re-written in this form in three alternative ways: the choice \( \gamma = c \) gives \( \varepsilon((a, c)) = \varepsilon((b, c)) - \varepsilon((a, b, c)) \leq \varepsilon((a)) + \varepsilon((b)) - \varepsilon((a, b)) \) because of the convention \( \varepsilon([a]) = 1 = \varepsilon([b]) \).

4. Clutter inequalities

A deeper observation derived from our analysis of the cases \( n = |N| \leq 5 \) is that the discussed inequalities can be interpreted as inequalities induced by certain clutters of subsets of \( N \). These special clutter inequalities appear to define facets of the chordal graph polytope in general, that is, for any \( |N| \) (see Sections 7 and 9). Now, we introduce those inequalities formally.
**Definition 3 (clutter inequality).** Let \( \mathcal{L} \) be a clutter of subsets of \( N \) satisfying \( \emptyset \neq \bigcup \mathcal{L} \). The clutter inequality induced by \( \mathcal{L} \) is a linear constraint on \( c \in \mathbb{R}^{2^N} \) of the form

\[
1 \leq v(c, \mathcal{L}) := \sum_{S \subseteq N} \left( \kappa_L(S) \cdot c(S) \right) = \sum_{S \subseteq N} \left( \sum_{T \subseteq S} \kappa_L(T) \right) \cdot c(S) \cdot \sum_{T \subseteq S} (-1)^{|\mathcal{L} \setminus T|} \cdot \kappa_L(T) \cdot c(S) \cdot \sum_{T \subseteq S} (-1)^{|\mathcal{L} \setminus T|} \cdot \kappa_L(T) \cdot c(S)
\]  

(1)

In this context, recall the convention \( c(L) = 1 \) for any \( L \subseteq N, |L| = 1 \). We have formally introduced the inequality for any non-trivial clutter \( \mathcal{L} \), which appears to be convenient. Nonetheless, we are going to show in Section 7 that (1) follows from above conventional equality constraints for any \( c \in D_N \) only when \( \mathcal{L} \) contains a singleton. If \( \mathcal{L} \) consists of a sole singleton then (1) follows from above conventional equality constraints \( c([i]) = 1 \) for \( i \in N \).

One can write the clutter inequality in various forms. In this section we describe a simple way to compute the coefficients with sets in (1), give its unique form in the space \( \mathbb{R}^T \) for the polytope \( D_N \) and explain its generalized monotonicity interpretation. Later, in Section 6 we re-write the clutter inequality (1) in terms of other vector representatives of chordal graphs.

**Lemma 1 (basic re-writings of the clutter inequality).** Let \( \mathcal{L} \) be a clutter of subsets of \( N \) such that \( \emptyset \neq \bigcup \mathcal{L} \). Given \( c \in \mathbb{R}^{2^N} \), the value \( v(c, \mathcal{L}) \) from (1) can be expressed as follows:

\[
1 \leq v(c, \mathcal{L}) = \sum_{S \subseteq N} \kappa_L(S) \cdot c(S) \quad \text{where} \quad \kappa_L(S) := \sum_{0 \neq \mathcal{B} \subseteq \mathcal{L} : \mathcal{B} \ni S} (-1)^{|\mathcal{B}|+1} \quad \text{for any} \quad S \subseteq N.
\]  

(2)

The coefficients \( \kappa_L(\cdot) \) in (2) vanish outside the set system

\[
U(\mathcal{L}) := \left\{ \bigcup B : \emptyset \neq B \subseteq \mathcal{L} \right\}
\]

of unions of sets from \( \mathcal{L} \).

Within this set system, they can be computed recursively using the formula

\[
\kappa_L(S) = 1 - \sum_{T \in U(\mathcal{L}) : T \subseteq S} \kappa_L(T) \quad \text{for any} \quad S \in U(\mathcal{L}),
\]  

(3)

which implicitly says that \( \kappa_L(L) = 1 \) for \( L \in \mathcal{L} \). The formula (2) gets its unique form

\[
1 - |\mathcal{L} \setminus T| \leq \sum_{S \subseteq T} \kappa_L(S) \cdot c(S)
\]  

(4)

in the linear space \( \mathbb{R}^T \), where the polytope \( D_N \) is full-dimensional. For this reason, we call \( \mathbb{R}^T \) the *proper linear space* and refer to (4) as the *proper-space form* of the inequality. Finally, the coefficient vector \( \kappa_L \in \mathbb{R}^{2^N} \) from (2) is closely related to the indicator of \( \mathcal{L} \), the filter generated by \( \mathcal{L} \):

\[
\delta(T \in \mathcal{L}) := \sum_{S \subseteq T} \kappa_L(S) \quad \text{for any} \quad T \subseteq N.
\]  

(5)

**Proof.** We re-arrange the terms in (1) after the sets \( S = \bigcup \mathcal{B} \) and get

\[
v(c, \mathcal{L}) = \sum_{0 \neq \mathcal{B} \subseteq \mathcal{L}} (-1)^{|\mathcal{B}|+1} \cdot c(\bigcup \mathcal{B}) = \sum_{S \subseteq N} c(S) \cdot \sum_{0 \neq \mathcal{B} \subseteq \mathcal{L} : \mathcal{B} \ni S} (-1)^{|\mathcal{B}|+1} \cdot \kappa_L(S)
\]

which gives (2). It is immediate from this that \( \kappa_L(S) = 0 \) once \( S \notin U(\mathcal{L}) \). Having fixed \( S \in U(\mathcal{L}) \) observe that the set system \( \mathcal{L}_S := \{ L \in \mathcal{L} : L \subseteq S \} \) is non-empty, which allows us to write:

\[
\sum_{T \in U(\mathcal{L}) : T \subseteq S} \kappa_L(T) = \sum_{T \subseteq S} \kappa_L(T) = \sum_{0 \neq \mathcal{B} \subseteq \mathcal{L} : \mathcal{B} \ni T} (-1)^{|\mathcal{B}|+1} = \sum_{0 \neq \mathcal{B} \subseteq \mathcal{L} : \mathcal{B} \ni S} (-1)^{|\mathcal{B}|+1} = 1 + \sum_{0 \neq \mathcal{B} \subseteq \mathcal{L} : \mathcal{B} \ni S} (-1)^{|\mathcal{B}|+1} = 1 - \sum_{0 \neq \mathcal{B} \subseteq \mathcal{L} : \mathcal{B} \ni S} (-1)^{|\mathcal{B}|} = 1.
\]
which gives \( \text{(3)} \). To transform \( \text{(2)} \) into \( \text{(4)} \) note that \( \kappa_L(0) = 0 \) and, for \( L \subseteq N, |L| = 1 \), one has \( c(L) = 1 \) while \( \kappa_L(L) \) is either 1 or 0, depending on whether \( L \in \mathcal{L} \) or not. Of course, the number of singletons in \( \mathcal{L} \) is just \( |\mathcal{L} \setminus \Upsilon| \).

To verify \( \text{(5)} \) fix \( T \subseteq N \), denote again \( L_T := \{ L \in \mathcal{L} : L \subseteq T \} \) and write analogously

\[
\sum_{S \subseteq T} \kappa_L(S) \leq \sum_{S \subseteq T} \sum_{\emptyset \neq B \subseteq L \setminus \Upsilon} (-1)^{|B|+1} = \sum_{\emptyset \neq B \subseteq L \setminus \Upsilon} (-1)^{|B|+1} = 1 + \sum_{\emptyset \neq B \subseteq L \setminus \Upsilon} (-1)^{|B|+1}
\]

and it remains to realize that \( L_T \neq \emptyset \) iff \( T \in \mathcal{L}' \).

Let us illustrate Lemma \( \text{(1)} \) by an example; it hopefully indicates that, for small clutters \( \mathcal{L} \), the respective coefficient vector \( \kappa_L \in \mathbb{R}^T \) in the proper-space form \( \text{(4)} \) of the inequality is sparse because \( |\mathcal{U}(\mathcal{L})| \) is small.

Example 1 (computing coefficients in a clutter inequality). Take \( N = \{a, b, c, d\} \) and \( \mathcal{L} = \{ \{a, b\}, \{a, c\}, \{b, c\}, \{d\} \} \). Then the fact that \( \kappa_L(L) = 1 \) for \( L \in \mathcal{L} \) gives \( \kappa_L([a, b]) = \kappa_L([a, c]) = \kappa_L([b, c]) = \kappa_L([d]) = 1 \). The remaining elements in \( \mathcal{U}(\mathcal{L}) \) are \( \{a, b, d\}, \{a, c, d\}, \{b, c, d\} \). The recursive formula \( \text{(3)} \) can be applied to \( \{a, b, d\} \), whose proper subsets in \( \mathcal{U}(\mathcal{L}) \) are \( \{a, b\} \) and \( \{d\} \), which yields

\[
\kappa_L([a, b, d]) = 1 - \kappa_L([a, b]) - \kappa_L([d]) = 1 - 1 - 1 = -1.
\]

Analogously, \( \kappa_L([a, c, d]) = \kappa_L([b, c, d]) = -1 \). As concerns \( \{a, b, c\} \), it has three proper subsets in \( \mathcal{U}(\mathcal{L}) \), which gives

\[
\kappa_L([a, b, c]) = 1 - \kappa_L([a, b]) - \kappa_L([a, c]) - \kappa_L([b, c]) = 1 - 1 - 1 - 1 = -2.
\]

Finally, \( \{a, b, c, d\} \) has all other sets in \( \mathcal{U}(\mathcal{L}) \) as proper subsets which gives

\[
\kappa_L([a, b, c, d]) = 1 - \kappa_L([a, b]) - \kappa_L([a, c]) - \kappa_L([b, c]) - \kappa_L([d]) - \kappa_L([a, b, d]) - \kappa_L([a, c, d]) - \kappa_L([b, c, d]) - \kappa_L([a, b, c]) - \kappa_L([a, c, b]) - \kappa_L([b, c, a]) - \kappa_L([a, b, c, d]) = 1 - 1 - 1 - 1 - 1 - 1 - 1 - 2 = 2
\]

Because the remaining coefficients \( \kappa_L(S) \) vanish and \( \mathcal{L} \) contains one singleton, that is, \( |\mathcal{L} \setminus \Upsilon| = 1 \), the proper-space formula \( \text{(4)} \) takes the form

\[
0 = 1 - |\mathcal{L} \setminus \Upsilon| \leq c([a, b]) + c([a, c]) + c([b, c]) - c([a, b, d]) - c([a, c, d]) - c([b, c, d]) - 2 \cdot c([a, b, c]) + 2 \cdot c([a, b, c, d]) - 2 \cdot c([a, b, c, d])
\]

4.1. Generalized monotonicity interpretation of clutter inequalities

Another interesting observation is that if a clutter contains a singleton then the corresponding inequality can be interpreted as a generalized monotonicity constraint. Indeed, given a clutter \( \mathcal{L} \subseteq \mathcal{P}(N) \) containing a singleton \( \{\gamma\} \) such that \( |\mathcal{L}| \geq 2 \), let us put \( \mathcal{R} := \mathcal{L} \setminus \{\gamma\} \). Then \( \bigcup \mathcal{R} \neq \emptyset \) and the formulas \( \text{(2)} \) and \( \text{(3)} \) from Lemma \( \text{(1)} \) allow one to observe that

\[
\kappa_L(S) = \begin{cases} 
\kappa_R(S) & \text{for } S \subseteq N \setminus \{\gamma\}, \\
-\kappa_R(S \setminus \{\gamma\}) & \text{for } S \subseteq N \setminus \{\gamma\}, S \setminus \{\gamma\} \neq \emptyset, \\
1 & \text{for } S = \{\gamma\}.
\end{cases}
\]

Because of the convention \( c(\{\gamma\}) = 1 \) and the fact \( \kappa_L(\emptyset) = 0 \) the formula \( \text{(2)} \) can be re-arranged into the following generalized monotonicity form:

\[
\sum_{S \subseteq N \setminus \{\gamma\}} \kappa_R(S) \cdot c(S \cup \{\gamma\}) \leq \sum_{S \subseteq N \setminus \{\gamma\}} \kappa_R(S) \cdot c(S).
\]

Observe that if \( \mathcal{L} \) contains several singletons then \( \text{(2)} \) also has several generalized monotonicity re-writings. Let us illustrate the formula \( \text{(7)} \) by an example.
Example 2 (generalized monotonicity form of a clutter inequality). Consider the same set \( N = \{a, b, c, d\} \) and clutter \( \mathcal{L} = \{\{a, b\}, \{a, c\}, \{b, c\}, \{d\}\} \) as in Example 1. Then necessarily \( \gamma = d \) and \( \mathcal{K} = \{\{a, b\}, \{a, c\}, \{b, c\}\} \) which leads to \( \kappa_{\mathcal{K}}(\{a, b\}) = \kappa_{\mathcal{K}}(\{a, c\}) = \kappa_{\mathcal{K}}(\{b, c\}) = 1 \) and \( \kappa_{\mathcal{K}}(\{a, b, c\}) = -2 \). Thus, the generalized monotonicity form \( (7) \) is

\[
\kappa(a, b, d) + \kappa(a, c, d) + \kappa(b, c, d) - 2 \cdot \kappa(a, b, c, d) \leq \kappa(a, b) + \kappa(a, c) + \kappa(b, c) - 2 \cdot \kappa(a, b, c),
\]

which is just a re-writing of the inequality from Example 1.

5. Completeness conjecture

We have the following conjecture we know is valid in the case \(|N| \leq 5\).

Conjecture 1. For any \( n = |N| \geq 2 \), all facet-defining inequalities for \( c \in D_N \) are the lower bound \( 0 \leq \kappa(N) \) and the inequalities \( (1) \) induced by clutters \( \mathcal{L} \) of subsets of \( N \) that contain at least one singleton and satisfy \( |\bigcup \mathcal{L}| \geq 2 \).

Recall that the convention \( \kappa(L) = 1 \) for \( L \subseteq N \), \( |L| = 1 \), implies that \( (1) \) holds with equality provided \( |\bigcup \mathcal{L}| = 1 \). On the other hand, if a clutter \( \mathcal{L} \) with \( |\bigcup \mathcal{L}| \geq 2 \) does not contain a singleton then \( (1) \) is not valid for \( c \in D_N \) since the characteristic inset of the empty graph produces the RHS of zero in \( (1) \).

Conjecture 1 can be viewed as a substantial step towards the solution to a more general problem when a prescribed clique size limit is given.

Conjecture 2. For any \( 2 \leq k \leq n \), a polyhedral description of \( D_N^k \) is given by

- the lower bounds \( 0 \leq \kappa(K) \) for \( K \subseteq N \), \( |K| = k \), and
- the inequalities \( (1) \) induced by clutters \( \mathcal{L} \) which are subsets of the set system \( \{L \subseteq N : |L| < k\} \), contain at least one singleton and satisfy \( |\bigcup \mathcal{L}| \geq 2 \).

We will shown in Example 4 from Section 3 that not every inequality from Conjecture 2 is facet-defining for \( D_N^k \); thus, the problem of exact characterization of facets of \( D_N^k \) is more subtle.

5.1. Cluster inequalities in the case of 4 variables

To illustrate Conjecture 1, let us list the 9 types of the 50 facet-defining inequalities for \( D_N \) in case \( n = |N| = 4 \) and interpret them in terms of clutters. An exceptional case, which is not a clutter inequality, is the lower bound:

- **lower bound:** \( 0 \leq \kappa(a, b, c, d) \) (1 inequality).

Two types of monotonicity inequalities correspond to quite simple clutters, namely to one singleton together with one non-singleton:

- **3-to-4 monotonicity:** take \( \mathcal{L} = \{\{a, b, c\}, \{d\}\} \), \( (2) \) gives \( 1 \leq \kappa(a, b, c) + \kappa(d) - \kappa(a, b, c, d) \) and, because of \( \kappa(d) = 1 \), one gets \( \kappa(a, b, c, d) \leq \kappa(a, b, c) \) (4 inequalities).

- **2-to-3 monotonicity:** take \( \mathcal{L} = \{\{a, b\}, \{c\}\} \), \( (3) \) gives \( 1 \leq \kappa(a, b) + \kappa(c) - \kappa(a, b, c) \) and, because of \( \kappa(c) = 1 \), one gets \( \kappa(a, b, c) \leq \kappa(a, b) \) (12 inequalities).

The cluster inequalities, whose special cases are the upper bounds, correspond to clutters consisting of singletons only (see Example 3 for details):

- **upper bounds:** take \( \mathcal{L} = \{\{a\}, \{b\}\} \), \( (4) \) gives \( 1 \leq \kappa(a) + \kappa(b) - \kappa(a, b) \) and, since \( \kappa(a) = \kappa(b) = 1 \), one gets \( \kappa(a, b) \leq 1 \) (6 inequalities).

- **cluster for 3-element-sets:** take \( \mathcal{L} = \{\{a\}, \{b\}, \{c\}\} \), \( (5) \) gives

\[
1 \leq \kappa(a) + \kappa(b) + \kappa(c) - \kappa(a, b) - \kappa(a, c) - \kappa(b, c) + \kappa(a, b, c) \quad \text{and one gets}
\]

\[
\kappa(a, b) + \kappa(a, c) + \kappa(b, c) \leq 2 + \kappa(a, b, c) \quad \text{(4 inequalities)},
\]
6.1. Clutter inequalities in terms of Möbius inversion

Moreover, we re-write (1) in terms of family variable vectors, which is equivalent to the condition

\[ c(a, b) + c(a, c) + c(a, d) + c(b, c) + c(b, d) + c(c, d) + c[(a, b, c, d)] \leq 3 + c(a, b, c) + c(a, b, d) + c(a, c, d) + c(b, c, d) \quad (1 \text{ inequality}). \]

Besides 28 “basic” inequalities, which have already occurred in the case \( n = 3 \), there are additionally 22 non-basic inequalities decomposing into 3 types; we gave them some auxiliary labels:

- **one 2-element-set clutter**: take \( \mathcal{L} = \{ [a, b], [c], [d] \} \) and (\( \square \)) leads to
  \[
  c(a, c) + c(a, d) + c(b, c) \leq 1 + c(a, b) + c((a, b, c, d)) \quad (6 \text{ inequalities}),
  \]

- **two 2-element-set clutter**: take \( \mathcal{L} = \{ [a, c], [b, c], [d] \} \) and (\( \square \)) leads to
  \[
  c(a, b, c) + c(a, c, d) + c(b, c, d) \leq c(a, c) + c(b, c) + c((a, b, c, d)) \quad (12 \text{ inequalities}),
  \]

- **three 2-element-sets clutter**: take \( \mathcal{L} = \{ [a, c], [a, c], [b, c], [d] \} \) and (\( \square \)) leads to
  \[
  2 \cdot c(a, b, c) + c(a, b, d) + c(a, c, d) + c(b, c, d) \leq c(a, b) + c(a, c) + c(b, c) + 2 \cdot c(a, b, c, d) \quad (4 \text{ inequalities}).
  \]

Note that the last inequality is equivalent to the one from Examples 1 and 2.

In the case \( n = |N| = 4 \) and the clique size limit \( k = 3 \), the restricted polytope \( D^3_N \) has 60 vertices, encoding 60 chordal graphs in which \( N \) is not complete. The polytope is specified by 1 equality constraint and 49 facet-defining inequalities, decomposing into 8 permutation types. These are obtained from the above ones by the substitution \( c((a, b, c, d)) = 0 \). Thus, the number of facets is nearly the same as in the unrestricted case.

However, the polytope \( D^2_N \) with \( |N| = 4 \) and \( k = 2 \) is considerably simpler: it has 38 vertices, encoding 38 undirected forests over four nodes. The polytope is specified by 5 equality constraints of the form \( c((a, b, c, d)) = 0 = c((a, b, c)) \), and by 17 facet-defining inequalities decomposing into 4 permutation types. These are either the lower bounds of the form \( 0 \leq c((a, b)) \) or the cluster inequalities of 3 types, including the upper bounds \( c((a, b)) \leq 1 \). In particular, some of the cluster inequalities mentioned above are not facet-defining in this subcase.

6. Other versions of clutter inequalities

To prove the validity of the clutter inequalities from Conjecture 1, it is useful to re-write them in terms of alternative vector representatives. In this section, we apply a convenient linear transformation to the vectors \( c \in \mathbb{R}^{P(N)} \) in (\( \square \)). Moreover, we re-write (\( \square \)) in terms of family variable vectors.

### 6.1. Clutter inequalities in terms of Möbius inversion

A very useful re-writing of the clutter inequality (\( \square \)) is in terms of a linear transformation of the vector \( c \in \mathbb{R}^{P(N)} \), known as the Möbius inversion.

**Definition 4** (superset Möbius inversion).

*Given a vector \( c \in \mathbb{R}^{P(N)} \), the superset Möbius inversion of \( c \) is the vector \( m \in \mathbb{R}^{P(N)} \) determined by the formula*

\[
m(T) := \sum_{S: T \subseteq S} (-1)^{|S|(T)} \cdot c(S) \quad \text{for any } T \subseteq N,
\]

*which is equivalent to the condition*

\[
c(S) = \sum_{T: S \subseteq T} m(T) \quad \text{for any } S \subseteq N.
\]
Indeed, to verify \((8)\Rightarrow(9)\) write for a fixed \(S \subseteq N:\)

\[
\sum_{T \subseteq N} m(T) \stackrel{(8)}{=} \sum_{T \subseteq N} \sum_{L \subseteq T} (-1)^{|L \setminus T|} \cdot c(L) = \sum_{L \subseteq N} c(L) \cdot \sum_{T \subseteq L} (-1)^{|L \setminus T|}
\]

\[
= \sum_{L \subseteq N} c(L) \cdot \sum_{B \subseteq L \setminus S} (-1)^{|B|} = \sum_{L \subseteq N} c(L) \cdot \delta(L \setminus S = \emptyset) = c(S).
\]

The proof of the implication \((9)\Rightarrow(8)\) is analogous.

Now, we give the form of clutter inequalities in this context. Let us note that the transformed coefficient vector need not be sparse even for small clutters \(\mathcal{L}\).

**Lemma 2** (clutter inequality in terms of superset M"obius inversion).

Let \(\mathcal{L}\) be a clutter of subsets of \(N\) such that \(\emptyset \neq \bigcup \mathcal{L}\). Then the clutter inequality induced by \(\mathcal{L}\) has the following form in terms of superset M"obius inversion \(m\) of the vector \(c \in \mathbb{R}^{\binom{N}{2}}\):

\[
1 \leq v(c, \mathcal{L}) = \sum_{T \subseteq N} \delta(T \in \mathcal{L}^\uparrow) \cdot m(T). \tag{10}
\]

Moreover, the formula \((10)\) has the following proper-space form

\[
1 - |\mathcal{L} \setminus \mathcal{T}| \leq \sum_{T \in \mathcal{T}} \lambda_\mathcal{L}(T) \cdot m(T), \quad \text{where } \lambda_\mathcal{L}(T) := \delta(T \in \mathcal{L}^\uparrow) - \sum_{i \notin T} \delta([i] \in \mathcal{L}) \quad \text{for any } T \in \mathcal{T}. \tag{11}
\]

in the linear space \(\mathbb{R}^\mathcal{T}\), where the respective polytope is full-dimensional.

The proof of Lemma 2 is given in Appendix A. Let us note (without a proof) that the relation of the coefficients in \((4)\) and in \((11)\) is that \(\kappa_\mathcal{L}\) is the subset M"obius inversion of \(\lambda_\mathcal{L}\) restricted to \(\mathcal{T}\):

\[
\lambda_\mathcal{L}(T) = \sum_{S \in \mathcal{T}, \delta S \leq T} \kappa_\mathcal{L}(S) \quad \text{for } T \in \mathcal{T}, \quad \text{and conversely}
\]

\[
\kappa_\mathcal{L}(S) = \sum_{T \in \mathcal{T}, \delta T \geq S} (-1)^{|S \setminus T|} \cdot \lambda_\mathcal{L}(T) \quad \text{for } S \in \mathcal{T}.
\]

The superset M"obius inversion \(m_\mathcal{L}\) of the characteristic ismet \(c_\mathcal{L}\) of a chordal graph \(G\) can serve as an alternative vector representative of the respective decomposable model. Here is the formula for \(m_\mathcal{L}\).

**Lemma 3** (superset M"obius inversion of the characteristic ismet).

Given a chordal graph \(G\) over \(N\), let \(m_\mathcal{L}\) denote the superset M"obius inversion of its characteristic ismet \(c_\mathcal{L}\), given by \((8)\) where \(c = c_\mathcal{L}\) and the convention \(c_\mathcal{L}(\emptyset) := 1\) is accepted. Assume that \(C(G)\) is the collection of cliques of \(G\), \(S(G)\) the collection of separators in \(G\) and let \(w_\mathcal{L}(S)\) denote the multiplicity of a separator \(S \in S(G)\). Then, for any \(T \subseteq N\),

\[
m_\mathcal{L}(T) = \sum_{C \in C(G)} \delta(T = C) - \sum_{S \in S(G)} w_\mathcal{L}(S) \cdot \delta(T = S) = \sum_{j=1}^m \delta(T = C_j) - \sum_{j=2}^m \delta(T = S_j), \tag{12}
\]

where \(C_1, \ldots, C_m, m \geq 1\), is an arbitrary ordering of elements of \(C(G)\) satisfying the RIP.

The proof of Lemma 3 can be found in Appendix B. It follows from the formula \((12)\) that \(m_\mathcal{L}\) need not be a zero-one vector because of multiplicities of separators. Nevertheless, in comparison with \(c_\mathcal{L}\), its superset M"obius inversion \(m_\mathcal{L}\) is typically a sparse vector in the sense that most of its components are zeros. The vector \(m_\mathcal{L}\) is a minor modification of the concept of a standard ismet treated already in [24] Section 7.2.2 and it is also close to zero-one encodings of junction trees used in [22].
6.2. Family variable formulation of clutter inequalities

This subsection requires a reader familiar with details of the ILP approach to learning Bayesian network structure. Recall from [7] that the family variable vector encoding a directed acyclic graph \( H \) over \( N \) is a zero-one vector \( \eta \) with components indexed by pair \((a, B)\), where \( a \in N \) and \( B \subseteq N \setminus \{a\} \); let us denote the component of \( \eta \) indexed by such a pair by \( \eta_{a\rightarrow B} \). Specifically, \( \eta_{a\rightarrow B} = 1 \) iff \( B = \text{pa}_a(a) \) is the set of parents of the node \( a \) in \( H \). Thus, every such vector belongs to the polyhedron of vectors \( \eta \) specified by the constraints \( 0 \leq \eta_{a\rightarrow B} \leq 1 \) for any \((a, B)\) and \( \sum_{B \subseteq N \setminus \{a\}} \eta_{a\rightarrow B} = 1 \) for any \( a \in N \), which is a common frame for family variable representatives.

Another possible (non-unique) vector representative of the decomposable model induced by a chordal graph \( G \) over \( N \) is any family variable vector \( \eta \) encoding a directed acyclic graph \( H \) over \( N \) inducing the same structural model as \( G \). There is a linear relation between the characteristic imset \( c = c_G \) and the family variable vector \( \eta \). Specifically, it was shown in [27, Lemma 3] that

\[
\eta(B) = \sum_{S \subseteq \{a\} \cup B} \eta(B \setminus a) \quad \text{for any } (a, B) \in \mathcal{L}.
\]

(13)

Recall in this context that the value \( c(\emptyset) \) for the empty set is irrelevant in (1). The formula (13) allows us to re-formulate the clutter inequality (1) in terms of family variables with zero-one coefficients.

**Lemma 4** (clutter inequality in terms of family variable vectors).

Let \( \mathcal{L} \) be a clutter of subsets of \( N \) such that \( \emptyset \neq \bigcup \mathcal{L} \). Then (1), re-written in terms of the family variable vector \( \eta \) inducing \( c \) through (13), takes the form

\[
\sum_{S \subseteq \{a\} \cup B} \eta(B \setminus a) = 1 \quad \text{for any } (a, B) \in \mathcal{L}.
\]

(14)

The proof of Lemma 4 was shifted to Appendix C. Let us illustrate the result by an example.

**Example 3** (cluster inequalities). Given a cluster of variables \( C \subseteq N \), \( |C| \geq 2 \), consider the clutter \( \mathcal{L} = \{ [a] : a \in C \} \). Then, in (14), \( \rho_{a\rightarrow B} = 1 \) iff \( a \in C \) and \( B \cap C = \emptyset \). In particular, the corresponding clutter inequality has the form

\[
1 \leq \sum_{a \in C} \sum_{B \subseteq N \setminus C} \eta_{a\rightarrow B}
\]

in family variables. This is a well-known cluster inequality mentioned in [14] and [7]; its interpretation is that the cluster \( C \) must contain at least one node which has no parent node in \( C \). One can derive from (14) in Lemma 1 that it has the form

\[
1 - |C| \leq \sum_{S \subseteq Y : S \subseteq C} (-1)^{|S|+1} \cdot c(S),
\]

in terms of the characteristic imset, which also follows from [27, Lemma 7]. The cluster inequalities are known to be facet-defining for the family-variable polytope, defined as the convex hull of all family variable vectors encoding directed acyclic graphs over \( N \); this can be derived from [8, Corollary 4]. Special cases of the cluster inequalities are the upper bounds (see Section 5.1) where \( |C| = 2 \).

7. Validity of clutter inequalities

To show the validity of the clutter inequality (1) for every \( c \in D_N \) we use its re-writing (10) in terms of Möbius inversion from Lemma 2 and the formula (12) for the Möbius inversion of a characteristic imset from Lemma 5.

**Corollary 1.** Given a chordal graph \( G \) over \( N \), let \( C_1, \ldots, C_m, m \geq 1 \), be any ordering of elements of the collection \( \mathcal{C}(G) \) of (all) cliques of \( G \) satisfying the RIP. Given a clutter \( \mathcal{L} \) of subsets of \( N \) with \( \emptyset \neq \bigcup \mathcal{L} \) one has

\[
\nu(c_G, \mathcal{L}) = \sum_{j=1}^{m} \delta(C_j \in \mathcal{L}^\uparrow) - \sum_{j=2}^{m} \delta(S_j \in \mathcal{L}^\uparrow).
\]

(15)
Proof. We write using the formulas \((10)\) and \((12)\):
\[
\nu(c_G, \mathcal{L}) = \sum_{T \subseteq N} \delta(T \in \mathcal{L}^1) \cdot m_G(T) = \sum_{T \subseteq N} \delta(T \in \mathcal{L}^1) \cdot \left[ \sum_{j=1}^m \delta(T = C_j) - \sum_{j=2}^m \delta(T = S_j) \right]
\]
\[
= \sum_{j=1}^m \sum_{T \subseteq N} \delta(T = C_j) \cdot \delta(T \in \mathcal{L}^1) - \sum_{j=2}^m \sum_{T \subseteq N} \delta(T = S_j) \cdot \delta(T \in \mathcal{L}^1) = \sum_{j=1}^m \delta(C_j \in \mathcal{L}^1) - \sum_{j=2}^m \delta(S_j \in \mathcal{L}^1),
\]
which concludes the proof of \((15)\).

Now, the proof of the validity of \((1)\) is easy.

**Theorem 1** (validity of clutter inequalities).

Given a chordal graph \(G\) over \(N\), \(|N| \geq 2\), all inequalities from Conjecture \([1]\) are valid for the characteristic imset \(c_G\).

**Proof.** The validity of the lower bound \(0 \leq c_G(N)\) is immediate. As concerns \((1)\), given a clutter \(\mathcal{L}\) of subsets of \(N\) containing a singleton \(\{\gamma\}\), choose a clique \(C \in C(G)\) containing \(\gamma\) and an ordering \(C_1, \ldots, C_m, m \geq 1\), of cliques of \(G\) satisfying the RIP and \(C_1 = C\). Such an ordering exists by \([16, \text{Lemma 2.18}]\). By Corollary \([1]\) one has
\[
\nu(c_G, \mathcal{L}) = \sum_{T \subseteq N} \delta(C_1 \in \mathcal{L}^1) + \sum_{j=2}^m \left\{ \delta(C_j \in \mathcal{L}^1) - \delta(S_j \in \mathcal{L}^1) \right\} \geq 1,
\]
because \(\{\gamma\} \in \mathcal{L}\) implies \(C_1 \in \mathcal{L}^1\) and, also, \(S_j \in \mathcal{L}^1, S_j \subseteq C_j \Rightarrow C_j \in \mathcal{L}^1\).

\[\square\]

8. The clutter inequalities define facets

We observe that every inequality induced by a singleton-containing clutter is facet-defining for the unrestricted chordal graph polytope \(D_N\). In fact, we are going to show the next result in the case of a prescribed clique size limit.

**Lemma 5.** Given \(2 \leq k \leq n = |N|\), let \(\mathcal{L}\) be a clutter of subsets of \(N\) containing a singleton such that \(|\bigcup \mathcal{L}| \geq 2\) and \(|L \cup R| \leq k\) for any \(L, R \in \mathcal{L}\). Then the inequality \((1)\) induced by \(\mathcal{L}\) is facet-defining for \(D_N^k\).

Since the proof is very long it is shifted to Appendix D. Let us note that Lemma \([5]\) need not hold without the assumption \(|L \cup R| \leq k\) for \(L, R \in \mathcal{L}\) as Example \([4]\) below shows. On the other hand, it provides solely a sufficient condition on a clutter \(\mathcal{L}\) to define a facet of \(D_N^k\). In fact, the procedure from Appendix D allows one to verify the conclusion that the clutter inequality ascribed to a particular \(\mathcal{L}\) is facet-defining for \(D_N^k\) even in some rare cases when \(|L \cup R| > k\) for certain \(L, R \in \mathcal{L}\); as mentioned in Section \(5\) below Conjecture \([2]\), the task of characterizing facet-defining clutter inequalities for \(D_N^k, k > 2\), is open.

**Example 4** (non-facet clutter inequality in the restricted case). If \(n = |N| = 5\) and \(k = 3\) then consider the clutter \(\mathcal{L} = \{[a, b], [c, d], [e]\}\) with \(N = \{a, b, c, d, e\}\). Thus, \(\mathcal{L}\) is a subset of \([L \subseteq N : |L| < k]\) mentioned in Conjecture \([2]\), but the condition from Lemma \([5]\) is not fulfilled. By \((4)\), the proper-space version of the clutter inequality for \(\mathcal{L}\) has the next form in this restricted case:
\[
0 \leq \sigma([a, b]) + \sigma([c, d]) - \sigma([a, b, e]) - \sigma([c, d, e]) \quad \text{for } c \in D_5^3.
\]
This is, however, the sum of the inequalities
\[
0 \leq \sigma([a, b]) - \sigma([a, b, e]), \quad 0 \leq \sigma([c, d]) - \sigma([c, d, e]) \quad \text{for } c \in D_5^3,
\]
which are the clutter inequalities induced by \(L_1 = \{[a, b], [e]\}\) and by \(L_2 = \{[c, d], [e]\}\). To see that their sum is not facet-defining for \(D_5^3\), it is enough to observe that they define different proper faces of \(D_5^3\). To this end consider a (chordal) graph \(G\) with cliques \([c, d], [a]\) and \([b]\) and observe that \(c_G \in D_5^3\) is tight for the first inequality but not for the second one; conversely, \(c_H \in D_5^3\) where \(H\) has cliques \([a, b], [c]\) and \([d]\) is not tight for the first inequality (but it is tight for the second one).
Now, the main result follows.

**Theorem 2** (clutter inequalities define facets).
For every clutter $\mathcal{L} \subseteq \mathcal{P}(\mathcal{N})$ containing a singleton and satisfying $|\bigcup \mathcal{L}| \geq 2$, the corresponding inequality \(1\) is facet-defining for $D_N = D_N^\mathcal{R}$.

*Proof.* If $k = n$ then the condition on $\mathcal{L}$ from Lemma 5 is fulfilled. \(\square\)

### 9. The separation problem in the cutting plane method

The effort to find a complete polyhedral description of the polytope $D_N^\mathcal{R}$ from Section 2.5 is motivated by the aim to apply a *linear programming* (LP) approach to learning decomposable models. More specifically, as explained in Section 2, the statistical learning task can, in principle, be transformed into an LP problem to maximize a linear function over the (restricted) chordal graph polytope.

However, since every clutter inequality is facet-defining for $D_N$ (see Section 8), the number of inequalities describing $D_N$ is super-exponential in $n = |\mathcal{N}|$ and the use of a pure LP approach is not realistic. Instead, *integer linear programming* (ILP) methods can be applied, specifically the *cutting plane method* \(7\). In this approach, the initial task is to solve an LP problem which is a relaxation of the original problem, namely to maximize the objective over \(\mathcal{L} \subseteq \mathcal{P}(\mathcal{N})\) with \(|\bigcup \mathcal{L}| \geq 2\). There is a requirement that integer vectors in \(\mathcal{P}\) and $D_N$ coincide: $Z^\mathcal{R} \cap \mathcal{P} = Z^\mathcal{R} \cap D_N$. Moreover, facet-defining inequalities for $D_N$ appear to be the most useful ones, leading to good overall performance.

In this approach, if the optimal solution $c^*$ to the relaxed problem has only integer components then it is also the optimal solution to the unrelaxed problem. Otherwise, one has to solve the *separation problem* \(28\), which is to find a linear constraint (that is, a *cutting plane*) which separates $c^*$ from $D_N$. This new constraint is added and the method repeats starting from this new more tightly constrained problem. If our search is limited to the clutter inequalities then it leads to the following task:

- Given $c^* \notin D_N$, find clutter(s) $\mathcal{L}$ such that the inequality \(1\) is (most) violated by $c^*$, in other words, minimize $\mathcal{L} \mapsto v(c^*, \mathcal{L})$ over singleton-containing clutters $\mathcal{L} \subseteq \mathcal{P}(\mathcal{N})$ with $|\bigcup \mathcal{L}| \geq 2$.

Our idea is to re-formulate this task in the form of a few auxiliary LP problems. To this end we fix an element $\gamma \in \mathcal{N}$ and limit our search to clutters $\mathcal{L}$ with $\{\gamma\} \in \mathcal{L}$ and $(\bigcup \mathcal{L}) \setminus \{\gamma\} \neq \emptyset$. Thus, we decompose the whole separation problem into $n = |\mathcal{N}|$ subproblems.

To solve the subproblem with fixed $\gamma \in \mathcal{N}$ we denote

$$M := \mathcal{N} \setminus \{\gamma\}, \quad \mathcal{R} := \mathcal{L} \setminus \{\{\gamma\}\}$$

for any considered clutter $\mathcal{L}$

and realize that $\mathcal{R}$ is a clutter of subsets of $M$ with $\emptyset \neq \bigcup \mathcal{R}$. Write using the formulas from Section 4 and the convention $c^*(L) = 1$ for $L \subseteq N, |L| = 1$:

$$v(c^*, \mathcal{L}) = 1 + \sum_{S \subseteq \mathcal{N}} \kappa_L(S) \cdot c^*(S) - 1 - \sum_{S \subseteq \mathcal{M}} \kappa_R(S) \cdot c^*(S) - \sum_{\emptyset \neq L \subseteq \mathcal{M}} \kappa_R(L) \cdot c^*(L \cup \{\gamma\}) + c^*(\{\gamma\}) - 1$$

$$= \sum_{S \subseteq \mathcal{M}} \kappa_R(S) \cdot \left[ c^*(S) - c^*(S \cup \{\gamma\}) \right],$$

because of $\kappa_R(\emptyset) = 0$. Thus, the subproblem is to minimize

$$\mathcal{R} \mapsto \sum_{S \subseteq \mathcal{M}} \kappa_R(S) \cdot \left[ c^*(S) - c^*(S \cup \{\gamma\}) \right]$$

over clutters $\mathcal{R} \subseteq \mathcal{P}(M)$ with $\emptyset \neq \bigcup \mathcal{R}$ and this can be re-formulated in the form of an LP problem to minimize a linear objective $o^*$ over the clutter polytope

$$Q := \text{conv} \left( \{ \kappa_R \in \mathbb{R}^{\mathcal{P}(M)} : \mathcal{R} \subseteq \mathcal{P}(M) \text{ is a clutter with } \bigcup \mathcal{R} \neq \emptyset \} \right).$$

(17)
Note that the inequality \([1]\) corresponding to \(L = R \cup \{\emptyset\}\) is violated by \(c^*\) iff the respective value of the objective in \([16]\) is strictly negative. Moreover, provided the monotonicity inequalities (see Section 5.1) are involved in the specification of the starting relaxation \(P\) the objective vector \(\sigma^* \in \mathbb{R}^{|P(M)|}\) in \([16]\) has non-negative components. Below we give a polyhedral description of the clutter polytope \(Q\), which is surprisingly simple: if \(|M| \geq 3\) then the number of facets of \(Q\) is smaller than the number of its vertices.

The proof of our result is based on the following auxiliary observation; recall that a filter is a set system closed under supersets.

**Lemma 6** (polyhedral description of a transformed clutter polytope). Let \(M\) be a non-empty finite set. Given \(F \subseteq \mathcal{P}(M)\), introduce
\[
\sigma_F(T) := \delta(T \in F) \quad \text{for } T \subseteq M
\]
the indicator vector of \(F\). Then the filter polytope
\[
R := \text{conv}(\{\sigma_F \in \mathbb{R}^{\mathcal{P}(M)} : F \subseteq \mathcal{P}(M) \text{ is a filter with } \emptyset \notin F, M \in F \})
\]
is characterized by the following linear constraints:
\[
\sigma(\emptyset) = 0, \quad \sigma(M) = 1, \quad \sigma(B) \leq \sigma(B \cup \{a\}) \quad \text{for } a \in M, B \subseteq M \setminus \{a\}.
\]

The proof of Lemma 6 is in Appendix E. Now, one can show the following.

**Theorem 3** (polyhedral description of the clutter polytope). The clutter polytope \(Q\) from \([17]\) is determined by the following linear constraints on \(\kappa \in \mathbb{R}^{\mathcal{P}(M)}\):
\[
\begin{align*}
0 &= \kappa(\emptyset), \quad 1 = \sum_{S \subseteq M} \kappa(S), \\
0 &\leq \sum_{L \subseteq R} \kappa(L \cup \{a\}) \quad \text{for any pair } (a, B) \text{ where } a \in M, B \subseteq M \setminus \{a\}.
\end{align*}
\]

Observe that the inequalities from Theorem 3 imply \(0 \leq \kappa(\{a\})\) for any \(a \in M\). Let us note that the number of inequalities in Theorem 3 is just the number of family variables for \(M\), that is, \(|M| \cdot 2^{M-1}\), or equivalently, the number of edges of the Hasse diagram for the poset \((\mathcal{P}(M), \subseteq)\).

**Proof.** The idea is to use a suitable linear transformation. The formula \([5]\) in Lemma 1 means that \(\kappa_R\) is the subset Möbius inversion of the indicator of \(F := \mathcal{R}^I\), the filter generated by \(\mathcal{R}\), that is,
\[
\sigma_F(T) = \delta(T \in F) = \delta(T \in \mathcal{R}^I) = \sum_{S \subseteq T} \kappa_R(S) \quad \text{for any } T \subseteq M.
\]
The one-to-one linear mapping \(\kappa \leftrightarrow \sigma\) transforms \(Q\) to the polytope \(R\) defined by \([18]\). It follows from Lemma 6 that \(R\) is specified by constraints \([19]\), which turn into the constraints mentioned in Theorem 3 because of the transformation formula \(\sigma(T) = \sum_{S \subseteq T} \kappa(S)\) for \(T \subseteq N\).

**10. Preliminary computational experiments**

We have implemented some methods for solving the separation problem from Section 9 by extending the GOBNILP system \([17]\) for learning Bayesian networks. This was done by adding a constraint handler for chordal graph learning to the development version of GOBNILP which can be found at

GOBNILP already looks for the deepest cutting planes which are the clutter inequalities, that is, the clutter inequalities where all clutter members are singletons (see Example 3). Extending this to find the guaranteed best clutter cut for all possible clutters, for example by exploiting Theorem 3 has proved (so far) to be too slow. Instead preliminary results indicate that an approximate approach is superior: monotonicity inequalities \((|L| = 2)\) are added initially and then the separation problem is solved approximately by searching only for clutters where \(|L| \in \{3, 4\}\).
With this approach, the development version of GOBNILP can find an optimal chordal graph for the BRIDGES UCU dataset (12 variables, 108 datapoints) in 230s.

In contrast, as shown by Kangas et al. [15], the current stable version of GOBNILP, which learns chordal graphs by simply ruling out immoralities, cannot solve this problem even when given several hours.

Thus, the development version of GOBNILP is a clear improvement in comparison with the current version. However, when there is no limit on clique size, performance remains far behind that of the Junctor algorithm [15] which, for example, can solve the BRIDGES learning problem in only a few seconds.

Interestingly, with the separation algorithm turned off and no monotonicity inequalities added the development version of GOBNILP could still not solve this problem after 59,820s at which point we aborted since GOBNILP was using 12Gb of memory.

This shows the practical importance of using the clutter inequalities in an ILP approach to chordal graph learning.

Our conclusion from the preliminary empirical experiments is that the present poor performance is mainly caused by the large number of ILP variables one has to create. This is because one cannot apply the normal pruning for Bayesian network learning, as it has already been noted by Kangas et al. [15, §4]. Given our present state of knowledge, only when one restricts the maximal clique size (called traditionally treewidth) is there hope for reasonable performance. Thus, more extensive experimentation is delayed until further progress in pruning methods is achieved.

11. Conclusion: further theoretical results and open tasks

We have achieved several theoretical results on the clutter inequalities. In particular, we have succeeded in showing that every inequality from Conjecture 1 is facet-defining for the chordal graph polytope $D_N$.

There are further supporting arguments for the conjectures from Section 5. More specifically, we are able to show using a classic matroid theory result by Edmonds that a complete polyhedral description for $D^2_N$ consists of the lower bounds and the cluster inequalities. Thus, Conjecture 2 is true in case $k = 2$. We also have a promising ILP formulation for chordal graph learning using a subset of the facet-defining inequalities of $D_N$ as constraints. Nevertheless, to keep the length of this paper within standard limits we decided to postpone the proofs of these two results to a later publication.

The big theoretical challenge remains: to confirm/disprove Conjecture 1. Even if confirmed, another open problem is to characterize clutter inequalities defining facets for $D^k_N$, $2 \leq k \leq n$.

The preliminary empirical experiments indicate that a further theoretical goal should be to develop special pruning methods under the assumption that the optimal chordal graph is the learning goal. The result of such pruning procedure should be a confining system $\mathcal{K} \subseteq \mathcal{P}(N)$ of sets closed under subsets defining the general restricted chordal graph polytope (see Section 2.4). The subsequent goal, based on the result of pruning, can be to modify the proposed LP methods for solving the separation problem to become more efficient.

Acknowledgements

The research of Milan Studený has been supported by the grant GAČR number 16-12010S. We are grateful both to the reviewers of this journal paper and to the reviewers of our former PGM-2016 contribution for their comments. We particularly appreciate the detailed technical comments given by one of the journal paper reviewers.

Appendix A. Proof of Lemma 2

Let us recall what we are going to prove.

Rephrasing Lemma 2 Let $\mathcal{L}$ be a clutter of subsets of $N$ such that $\emptyset \not\in \bigcup \mathcal{L}$. Recall that the superset Möbius inversion $m$ of the vector $c \in \mathbb{R}^\mathcal{P}(N)$ satisfies

$$c(S) = \sum_{T: S \subseteq T} m(T) \quad \text{for any } S \subseteq N.$$
Then the clutter inequality \(1\) induced by \(\mathcal{L}\) has the following form in terms \(m\):

\[
1 \leq v(e, \mathcal{L}) = \sum_{T \subseteq N} \delta(T \in \mathcal{L}^1) \cdot m(T).
\]

Moreover, the formula \(10\) has the following unique form

\[
1 - |\mathcal{L} \setminus \mathcal{T}| \leq \sum_{T \in \mathcal{T}} A_{\mathcal{L}}(T) \cdot m(T), \text{ where } A_{\mathcal{L}}(T) := \delta(T \in \mathcal{L}^1) - \sum_{i \in T} \delta(|i| \in \mathcal{L}) \text{ for any } T \in \mathcal{T}.
\]

in the linear space \(\mathbb{R}^\mathcal{T}\).

Proof. Recall from Lemma \(1\) that the coefficient vector \(\kappa_{\mathcal{L}} \in \mathbb{R}^{\mathcal{P}(N)}\) in the basic version \(2\) of the clutter inequality \(1\) is related to the indicator of the corresponding filter \(\mathcal{L}^1\) by the formula

\[
\delta(T \in \mathcal{L}^1) = \sum_{S \subseteq T} \kappa_{\mathcal{L}}(S) \quad \text{for any } T \subseteq N.
\]

This allows us to write:

\[
v(e, \mathcal{L}) \in \sum_{S \subseteq N} \kappa_{\mathcal{L}}(S) \cdot \alpha(S) \in \sum_{S \subseteq N} \kappa_{\mathcal{L}}(S) \cdot \sum_{T : \mathcal{T} \subseteq \mathcal{S}} m(T) = \sum_{T \in \mathcal{T}} m(T) \cdot \sum_{S \subseteq N} \kappa_{\mathcal{L}}(S) \sum_{T \subseteq N} m(T) \cdot \delta(T \in \mathcal{L}^1),
\]

which concludes the proof of \(10\). To derive the formula \(11\) from \(10\) note that \(\emptyset \notin \mathcal{L}^1\); thus, for any \(i \in N\), one has \(\{i\} \in \mathcal{L}^1 \iff \{i\} \in \mathcal{L}\) and

\[
m(|i|) = \alpha(|i|) - \sum_{S \in \mathcal{T}, \{i\} \subseteq S} m(S) = 1 - \sum_{S \in \mathcal{T}, \{i\} \subseteq S} m(S),
\]

which allows one to write:

\[
v(e, \mathcal{L}) \in \sum_{T \subseteq \mathcal{T}} m(T) \cdot \delta(T \in \mathcal{L}^1) + \sum_{i \in N} m(|i|) \cdot \delta(|i| \in \mathcal{L})
\]

\[
= \sum_{T \subseteq \mathcal{T}} m(T) \cdot \delta(T \in \mathcal{L}^1) + \sum_{i \in N} \left[1 - \sum_{S \in \mathcal{T}, \{i\} \subseteq S} m(S)\right] \cdot \delta(|i| \in \mathcal{L})
\]

\[
= \sum_{i \in N} \delta(|i| \in \mathcal{L}) + \sum_{T \subseteq \mathcal{T}} m(T) \cdot \delta(T \in \mathcal{L}^1) - \sum_{i \in N} \sum_{S \in \mathcal{T}, \{i\} \subseteq S} m(S) \cdot \delta(|i| \in \mathcal{L})
\]

\[
= |\mathcal{L} \setminus \mathcal{T}| + \sum_{T \subseteq \mathcal{T}} m(T) \cdot \delta(T \in \mathcal{L}^1) - \sum_{i \in N} \sum_{S \in \mathcal{T}, \{i\} \subseteq S} \delta(|i| \in \mathcal{L})
\]

\[
= |\mathcal{L} \setminus \mathcal{T}| + \sum_{T \subseteq \mathcal{T}} m(T) \cdot \sum_{\lambda(T)} \delta(T \in \mathcal{L}^1) - \sum_{i \in N} \sum_{S \in \mathcal{T}, \{i\} \subseteq S} \delta(|i| \in \mathcal{L})
\]

which concludes the proof of \(11\). \(\square\)

Appendix B. Proof of Lemma 3

Let us recall what we are going to prove.

Recalling Lemma 3: Given a chordal graph \(G\) over \(N\), let \(m_G\) denote the superset Möbius inversion of its characteristic imset \(c_G\), where \(c = c_G\) and the convention \(c_G(\emptyset) = 1\) is accepted. Assume that \(C(G)\) is the collection of cliques of \(G\), \(S(G)\) the collection of separators in \(G\) and let \(w_G(S)\) denote the multiplicity of a separator \(S \in S(G)\). Then, for any \(T \subseteq N\),

\[
m_G(T) = \sum_{C \in C(G)} \delta(T = C) - \sum_{S \in S(G)} w_G(S) \cdot \delta(T = S) = \sum_{j=1}^{m} \delta(T = C_j) - \sum_{j=2}^{m} \delta(T = S_j).
\]

where \(C_1, \ldots, C_m, m \geq 1\), is an arbitrary ordering of elements of \(C(G)\) satisfying the RIP.
Proof. Let us put

\[ m'(T) := \sum_{B \subset C(G)} (-1)^{|B|+1} \cdot \delta(T = \bigcap B) \quad \text{for any } T \subseteq N; \]

the aim to show \( m' = m_G \). Thus, we denote

\[ C(G, S) := \{ C \in C(G) : S \subseteq C \} \quad \text{for any fixed } S \subseteq N, \]

and write

\[
\sum_{T \subseteq \mathcal{L}} m'(T) = \sum_{T \subseteq \mathcal{L}} \sum_{\emptyset \neq B \subset C(G)} (-1)^{|B|+1} \cdot \delta(T = \bigcap B) = \sum_{\emptyset \neq B \subset C(G)} (-1)^{|B|+1} \sum_{T \subseteq \mathcal{L}} \delta(T = \bigcap B) = \sum_{\emptyset \neq B \subset C(G)} (-1)^{|B|+1} = 1 + \sum_{\emptyset \neq B \subset C(G)} (-1)^{|B|+1} = \delta(C(G, S) \neq \emptyset) = \alpha_G(S).\]

Thus, \( \alpha_G \) is obtained from \( m' \) by the backward formula \(^9\). Hence, since the Möbius inversion is a one-to-one transformation, one has

\[
m_G(T) = \sum_{\emptyset \neq B \subset C(G)} (-1)^{|B|+1} \cdot \delta(T = \bigcap B) \quad \text{for any } T \subseteq N. \tag{B.1}
\]

The formula \((B.1)\) can be re-written: given any ordering \( C_1, \ldots, C_m, m \geq 1 \), of all cliques of \( G \) satisfying the RIP and the separators \( S_i = C_i \cap \bigcup_{j \neq i} C_j \), \( i = 2, \ldots, m \), one has

\[
m_G(T) = \delta(T = C_1) + \sum_{j=2}^m \{ \delta(T = C_j) - \delta(T = S_j) \} \quad \text{for } T \subseteq N. \tag{B.2}
\]

Indeed, \((B.2)\) can be derived from \((B.1)\) by induction on \( m \): if \( C = C_m, m \geq 2 \), then a preceding clique \( K = C_j, j < m \), exists with \( S_m = C \cap K \) and one has

\[
\sum_{B \subset C(G), C \in B} (-1)^{|B|+1} \cdot \delta(T = \bigcap B) = \delta(T = C) - \delta(T = C \cap K),
\]

because the other terms cancel each other (this follows from the RIP). The above formula then justifies the induction step because \( C(G) \setminus \{ C \} \) is also the collection of cliques of a chordal graph (over a smaller set of variables).

Since the order of cliques is irrelevant in \((B.1)\), the expression in \((B.2)\) does not depend on the choice of the ordering satisfying the RIP. In particular, \((B.2)\) can be written in the form \((13)\), where \( w_G(S) \) is the number of \( 2 \leq j \leq m \) with \( S = S_j \) for \( S \in \mathcal{S}(G) \), which is the multiplicity of the separator \( S \). \qed

Appendix C. Proof of Lemma \(^4\)

Let us recall what we are going to prove.

Rephrasing Lemma \(^4\): Let \( \mathcal{L} \) be a clutter of subsets of \( N \) such that \( \emptyset \neq \bigcup \mathcal{L} \). Recall the formula relating \( c \in \mathbb{R}^Y \) to the family variable vector \( \eta \):

\[
c(S) = \sum_{a < \mathcal{L}} \sum_{B \subset S \setminus \{a\}} \eta_{a-B} \quad \text{for } \emptyset \neq S \subseteq N. \tag{13}
\]

Then the clutter inequality \((1)\), re-written in terms of \( \eta \) takes the form

\[
1 \leq v(c, \mathcal{L}) = \sum_{a \in \bigcup \mathcal{L}} \sum_{B \subset N \setminus \{a\}} \rho_{a-B} \cdot \eta_{a-B}, \quad \text{where}
\]

\[
\rho_{a-B} = \begin{cases} 
1 & \text{if there exists } L \in \mathcal{L} \text{ with } L \subseteq B \cup \{a\} \text{ while there is no } R \in \mathcal{L} \text{ with } R \subseteq B, \\
0 & \text{otherwise}.
\end{cases}
\]
Proof. Let us substitute (13) into (1) and get

\[
1 \leq \sum_{a \not\in \mathcal{B} \subseteq \mathcal{L}} (-1)^{|\mathcal{B}|+1} \cdot c(\bigcup \mathcal{B})\quad \text{(13)} \sum_{a \not\in \mathcal{B} \subseteq \mathcal{L}} (-1)^{|\mathcal{B}|+1} \cdot \sum_{a \not\in \mathcal{B} \subseteq \mathcal{L}} \eta_{a-B} = \sum_{a \not\in \mathcal{B} \subseteq \mathcal{L}} \eta_{a-B} \cdot \sum_{a \not\in \mathcal{B} \subseteq \mathcal{L}} (-1)^{|\mathcal{B}|+1}.
\]

To derive a formula for \( \rho_{a-B} \), with fixed \( a \in \bigcup \mathcal{L} \) and \( B \subseteq N \setminus \{a\} \), we put

\[
\mathcal{L}[a \mapsto B] := \{ L \in \mathcal{L} : a \in L \text{ and } L \subseteq B \cup \{a\} \},
\]

\[
\mathcal{L}[B] := \{ R \in \mathcal{L} : R \subseteq B \}.
\]

Firstly, we show that \( \mathcal{L}[B] \neq \emptyset \Rightarrow \rho_{a-B} = 0 \). To this end choose and fix \( R \in \mathcal{L}[B] \) and realize that the condition \( a \in \bigcup \mathcal{B} \subseteq B \cup \{a\} \) holds for \( \mathcal{B} \subseteq \mathcal{L} \) iff it holds for \( \mathcal{B} \cup \{R\} \), respectively for \( \mathcal{B} \setminus \{R\} \). Thus, the index set in the sum defining \( \rho_{a-B} \) decomposes into pairs \( \mathcal{B} \cup \{R\} \leftrightarrow \mathcal{B} \setminus \{R\} \) and one can write:

\[
\rho_{a-B} = \sum_{B \subseteq \mathcal{L}} (-1)^{|\mathcal{B}|+1} \sum_{a \not\in \mathcal{B} \subseteq \mathcal{L} \setminus \{a\}} (-1)^{|\mathcal{B}|+1} = \sum_{B \subseteq \mathcal{L}} (-1)^{|\mathcal{B}|+1} \sum_{a \not\in \mathcal{B} \subseteq \mathcal{L} \setminus \{a\}, R \in \mathcal{B}} (-1)^{|\mathcal{B} \cup \{R\}|+1} = \sum_{B \subseteq \mathcal{L}} (-1)^{|\mathcal{B}|+1} \sum_{a \not\in \mathcal{B} \subseteq \mathcal{L} \setminus \{a\}} (-1)^{|\mathcal{B} \cup \{R\}|+1} = 0.
\]

Secondly, assume that \( \mathcal{L}[B] = \emptyset \), that is, \( L \subseteq B \cup \{a\} \Rightarrow a \in L \) for any \( L \in \mathcal{L} \), and observe that in this case, for any \( B \subseteq \mathcal{L} \), one has

\[
\{ a \in \bigcup \mathcal{B} \subseteq B \cup \{a\} \} \Leftrightarrow \emptyset \neq B \subseteq \mathcal{L}[a \mapsto B].
\]

This allows one to write in the case \( \mathcal{L}[B] = \emptyset \):

\[
\rho_{a-B} = \sum_{B \subseteq \mathcal{L}} (-1)^{|\mathcal{B}|+1} = \sum_{a \not\in \mathcal{B} \subseteq \mathcal{L} \setminus \{a\}} (-1)^{|\mathcal{B}|+1} = 1 + \sum_{2 \subseteq \mathcal{L} \setminus \{a-B\}} (-1)^{|\mathcal{B}|+1} = \delta( \mathcal{L}[a \mapsto B] \neq \emptyset ).
\]

Hence, \( \rho_{a-B} = \delta( \mathcal{L}[B] = \emptyset ) \cdot \delta( \mathcal{L}[a \mapsto B] \neq \emptyset ) \), which gives (14) because in case \( \mathcal{L}[B] = \emptyset \) every \( L \in \mathcal{L} \), \( L \subseteq B \cup \{a\} \) contains \( \{a\} \) and belongs to \( \mathcal{L}[a \mapsto B] \).

Appendix D. Proof of Lemma 5

We base our proof on the following lemma, which is a kind of re-formulation of the method from [28, Approach 2 to Problem 1 in §9.2.3].

Lemma 7. Let \( \mathcal{P} \) be a full-dimensional polytope in \( \mathbb{R}^s \), \( s \geq 1 \), and

\[
\lambda_0 \leq \sum_{i=1}^s \lambda_i \cdot x_i \quad \text{for } x \equiv [x_1, \ldots, x_s] \in \mathbb{R}^s \quad (\text{where } \lambda_0, \lambda_1, \ldots, \lambda_s \in \mathbb{R}) \quad (D.1)
\]

a valid inequality for any \( x \in \mathcal{P} \), with at least one non-zero coefficient from \( \lambda_1, \ldots, \lambda_s \in \mathbb{R} \). Assume that there exist vectors \( x_1, \ldots, x_r \), \( r \geq s \), on the respective face, that is, vectors from \( \mathcal{P} \) satisfying (D.1) with equality, such that every real solution \( \mu_0, \mu_1, \ldots, \mu_r \) of the equations

\[
\forall j = 1, \ldots, r \quad \mu_0 = \sum_{i=1}^s \mu_i \cdot x_i^j \quad (D.2)
\]

is a multiple of \( \lambda_0, \lambda_1, \ldots, \lambda_s \), that is, \( \exists \alpha \in \mathbb{R} \quad \mu_i = \alpha \cdot \lambda_i \) for \( i = 0, 1, \ldots, s \).

Then the inequality (D.1) is facet-defining for \( \mathcal{P} \). In case \( r = s \) the vectors \( x_1, \ldots, x^s \), satisfying (D.2) are necessarily affinely independent.
Note that at least one of the coefficients \(\lambda_1, \ldots, \lambda_r \in \mathbb{R}\) is assumed to be non-zero since otherwise the existence of \(x^1, \ldots, x^r\) implies \(\lambda_0 = 0\) and \((D.1)\) is valid with equality for any \(x \in \mathbb{R}^s\) and, therefore, it is not facet-defining.

Proof. Firstly, observe that the condition \((D.2)\) implies that the affine hull of \(\{x^1, \ldots, x^r\}\) is an affine subspace of \(\mathbb{R}^s\) given by \(\lambda_0 = \langle \lambda, x \rangle := \sum_{i=1}^s \lambda_i \cdot x_i\).

Indeed, \(x \in \mathbb{R}^r\) belongs to the affine hull iff the corresponding extended vector \(\hat{x} := (1, x) \equiv (1, x_1, \ldots, x_r) \in \mathbb{R}^{r+1}\) is in the linear hull of extended vectors \(\hat{x}^1, \ldots, \hat{x}^s \in \mathbb{R}^{r+1}\): this is because for \(\beta_j \in \mathbb{R}, j = 1, \ldots, r\), one has

\[
(1, x) = \sum_{j=1}^r \beta_j \cdot (1, x^j) \quad \iff \quad \sum_{j=1}^r \beta_j = 1 \text{ and } x = \sum_{j=1}^r \beta_j \cdot x^j .
\]

Thus, it is enough to show that \((D.2)\) implies

\[
\text{Lin} (\{\hat{x}^1, \ldots, \hat{x}^s\}) = \{ (y_0, \ldots, y_s) \in \mathbb{R}^{r+1} : -\lambda_0 \cdot y_0 + \sum_{j=1}^s \lambda_j \cdot y_j = 0 \},
\]

where \(\text{Lin} (\cdot)\) denotes the linear hull and \(L\) the linear space specified by the constraint given by the coefficients \(\lambda_0, \lambda_1, \ldots, \lambda_r\); note that, for \(x \in \mathbb{R}^r\), one has \(\hat{x} = (1, x) \in L\) iff \(x\) satisfies \(\lambda_0 = \langle \lambda, x \rangle\).

The inclusion \(\text{Lin} (\{\hat{x}^1, \ldots, \hat{x}^s\}) \subseteq L\) is evident because vectors \(x^1, \ldots, x^r\) are assumed to belong to the face given by the respective inequality in \((D.1)\). The other inclusion \(L \subseteq \text{Lin} (\{\hat{x}^1, \ldots, \hat{x}^s\})\) is equivalent to the converse inclusion of their orthogonal complements \(\text{Lin} ((\hat{x}^1, \ldots, \hat{x}^s)^\perp) \subseteq L^\perp\). This is exactly what the condition \((D.2)\) requires: whenever \(\tilde{\mu} = (-\mu_0, \mu_1, \ldots, \mu_r) \in \text{Lin} ((\hat{x}^1, \ldots, \hat{x}^s)^\perp)\) then \(\tilde{\mu} \in \text{Lin} ((-\lambda_0, \lambda_1, \ldots, \lambda_r)) = L^\perp\).

Thus, provided \((D.2)\) holds, the affine hull of \(\{x^1, \ldots, x^r\}\) is determined by just one equality constraint in \(\mathbb{R}^s\) and has the dimension \(s - 1\), because \(\lambda_1, \ldots, \lambda_r\) are non-vanishing. In particular, the inequality \((D.1)\) defines a face of \(P\) of the dimension \(s - 1\), that is, a facet.

The conclusion that in case \(r = s\) the vectors \(x^1, \ldots, x^s\) satisfying \((D.2)\) are affinely independent can be derived as follows. In this case, the linear hull of \(\hat{x}^1, \ldots, \hat{x}^s \in \mathbb{R}^{r+1}\) is the space \(L\) of the dimension \(s\). But every set of \(s\) vectors linearly generating an \(s\)-dimensional subspace must be linearly independent. The linear independence of \(\hat{x}^1, \ldots, \hat{x}^s\) for \(\gamma_j \in \mathbb{R}, j = 1, \ldots, s\), leads to

\[
\sum_{j=1}^s \gamma_j = 0 \text{ and } \sum_{j=1}^s \gamma_j \cdot x^j = 0 \in \mathbb{R}^s \quad \Rightarrow \quad \sum_{j=1}^s \gamma_j \cdot \hat{x}^j = 0 \in \mathbb{R}^{r+1} \quad \Rightarrow \quad \gamma_j = 0 \text{ for } j = 1, \ldots, s ,
\]

that is, \(x^1, \ldots, x^s \in \mathbb{R}^s\) are affinely independent. \(\square\)

Let us recall Lemma 5 in more appropriate form before giving its proof.

**Rephrasing of Lemma 5.** Given \(2 \leq k \leq n = |N|\), let us denote

\[
\mathcal{K} := \{ S \subseteq N : |S| \leq k \}.
\]

Let \(\mathcal{L}\) be a clutter of subsets of \(N\) containing a singleton such that \(|\bigcup \mathcal{L}| \geq 2\) and \(L \cup R \in \mathcal{K}\) for any \(L, R \in \mathcal{L}\). Then the inequality \((1)\) induced by \(\mathcal{L}\) is facet-defining for \(D_N^k\).

**Proof.** The proof is more transparent if we transform the polytope \(D_N^k \subseteq D_N\) by the superset Möbius inversion \((8)\) and \(c \mapsto m\) to the polytope

\[
P := \text{conv} \{ m_G : G \text{ chordal graph over } N \text{ with clique size at most } k \}.
\]
and rewrite (1) accordingly. The dimension of $P$ is $\sum_{i=2}^k \binom{n}{i}$, the same like the one of $D_N^k$; the affine hull of $P$ is

$$A = \{ m \in \mathbb{R}^{P(N)} : m(T) = 0 \quad \text{for } T \notin \mathcal{K}, \quad \text{while} \quad \sum_{T \subseteq N} m(T) = 1 \quad \text{and} \quad \sum_{T \subseteq N : a \in T} m(T) = 1 \quad \text{for any } a \in N \},$$

where we use the fact that $P(N) \setminus \mathcal{K}$ is a filter. Elements of $P \subseteq A$ can be identified with vectors in $\mathbb{R}^{\mathcal{K} \cap \mathcal{T}}$ where

$$\mathcal{T} = \{ S \subseteq N : |S| \geq 2 \}$$

is the collection of non-empty non-singletons. This is because the restriction of $m \in A$ to components in $\mathcal{K} \cap \mathcal{T}$ determines affinely the values $m(T)$ for $T \subseteq N$ outside $\mathcal{K} \cap \mathcal{T}$. Moreover, $P$ is a full-dimensional polytope in $\mathbb{R}^{\mathcal{K} \cap \mathcal{T}}$, which fact can be derived from Lemma 3.

Given a singleton-containing clutter $L$ with $L, R \in \mathcal{L} \Rightarrow L \cup R \in \mathcal{K}$ and $|\cup \mathcal{L}| \geq 2$, we need appropriate rewriting of (1) in terms of $\mathbb{R}^{\mathcal{K} \cap \mathcal{T}}$. This is in Lemma 2, the formula (11), where we include the constraints for $m \in A$:

$$\lambda_0 \leq \sum_{T \in \mathcal{K} \cap \mathcal{T}} \lambda(T) \cdot m(T), \quad \text{for } m \in \mathbb{R}^{\mathcal{K} \cap \mathcal{T}} \quad \text{with} \quad \lambda_0 = 1 - |\mathcal{L} \setminus \mathcal{T}| \quad \lambda(T) = \delta(T \in \mathcal{L}^T) - \sum_{i \in T} \delta([i] \in \mathcal{L}) \quad \text{for } T \in \mathcal{K} \cap \mathcal{T}.

Observe that $|\cup \mathcal{L}| \geq 2$ implies that the coefficients in the RHS of (D.3) are not identically vanishing. One can derive from Theorem 1 that (D.3) is valid for any $m \in P$. Thus, we can use the criterion from Lemma 7 with $P \subseteq \mathbb{R}^{\mathcal{K} \cap \mathcal{T}}$ and the inequality (D.3).

To apply that criterion one has to construct a class $G$ of chordal graphs $G$ over $N$ with cliques in $\mathcal{K}$ that are tight for the clutter $L$, which means that $m_G$ satisfies (D.3) with equality. The vectors $m_G$ for $G \in \mathcal{G}$, viewed as elements in $\mathbb{R}^{\mathcal{K} \cap \mathcal{T}}$, will serve as the vectors on the face of $P$ given by (D.3); a formula for $m_G$ in $\mathbb{R}^{\mathcal{K} \cap \mathcal{T}}$ follows from (12) in Lemma 3

$$m_G(T) = \sum_{C \in \mathcal{G} \cap \mathcal{T}} \delta(T = C) - \sum_{S \in \mathcal{G} \cap \mathcal{T}} w_G(S) \cdot \delta(T = S) \quad \text{for } T \in \mathcal{K} \cap \mathcal{T}.

The goal is to construct such a class $G$ that the condition (D.2) from Lemma 7 holds for $[m_G : G \in \mathcal{G}]$ in place of $x^1, \ldots, x^r$, which means that, every collection of real numbers $\mu_0$ and $\mu(T), T \in \mathcal{K} \cap \mathcal{T}$, satisfying

$$\forall G \in \mathcal{G} \quad \mu_0 = \sum_{T \in \mathcal{K} \cap \mathcal{T}} \mu(T) \cdot m_G(T) \quad \text{(D.5)}$$

must be a multiple of the collection $\lambda_0$ and $\lambda(T), T \in \mathcal{K} \cap \mathcal{T}$, from (D.3). If we find such a class $G$ of graphs then Lemma 7 implies that (D.3) is facet-defining for $P$. The fact that the superset Möbius inversion (8) is linearly invertible by (9) then implies that (1) is facet-defining for $D_N^k$.

Roughly, a general principle of the construction of graphs as follows: for every $S \in \mathcal{K} \cap \mathcal{T}$, we include in $G$ a pair of graphs $G$ and $H$ such that the validity of (D.5) for $m_G$ and $m_H$ allows one to derive a conclusion on the value of $\mu(S)$. Given a clutter $\mathcal{L}$ containing a singleton and $|\cup \mathcal{L}| \geq 2$ we introduce

$$\Lambda := \bigcup_{i \in \mathcal{N} \setminus \{i \in \mathcal{L}\}} (\mathcal{L} \setminus \mathcal{T}) = \bigcup_{i \in \mathcal{N} \setminus \{i \in \mathcal{L}\}} [i], \quad \text{and} \quad \Gamma := N \setminus \Lambda,$$

and the details of the construction of graphs included in $G$ depend on whether

- $\lambda_0 < 0$, that is, $|A| \geq 2$, or
- $\lambda_0 = 0$, that is, $|A| = 1$, in other words, $\mathcal{L}$ only has one singleton.
The sets $S \in \mathcal{K} \cap \Gamma$ will be classified into 4 classes (that is, one has 4 cases of the construction):

A. if $|\Lambda| \geq 2$ sets $S \in \mathcal{K} \cap \Gamma$ such that $S \subseteq \Lambda$,

B. sets $S \in \mathcal{K} \cap \Gamma$ with $S \cap \Lambda \neq \emptyset \neq S \cap \Gamma$,

C. sets $S \in \mathcal{K} \cap \Gamma$ with $S \subseteq \Gamma$ and $S \notin \mathcal{L}$,

D. (if $\mathcal{L} \cap \Gamma \neq \emptyset$) sets $S \in \mathcal{K} \cap \Gamma$ with $S \subseteq \Gamma$ and $S \in \mathcal{L}$.

Moreover, one special graph will be constructed and included in $\mathcal{G}$ in order

E. to derive a conclusion on the constant $\mu_0$.

Now, the specific constructions in the above described cases will be given. Throughout the constructions, the vector $\delta_S \in \mathbb{R}^{\mathcal{K} \cap \Gamma}$, where $S \subseteq N$, will denote the zero-one identifier of the set $S$:

$$
\delta_S(T) := \begin{cases} 
1 & \text{if } T = S, \\
0 & \text{if } T \neq S, 
\end{cases} 
\text{ for } T \in \mathcal{K} \cap \Gamma.
$$

A. If $|\Lambda| \geq 2$ consider the collection of sets

$$
S := \{ S \in \mathcal{K} \cap \Gamma : S \subseteq \Lambda \},
$$

which is non-empty then. Every set $S \in S$ clearly belongs to $\mathcal{L}$ and, hence, one has $\lambda(S) = 1 - |S| < 0$ for such $S$. The whole consideration in this A-case has four steps. All these steps are empty in case $|\Lambda| = 2$ because then $|S| = 1$; thus, assume $|\Lambda| \geq 3$.

A.1. Verify that $\mu(S) = \mu(T)$ for every pair $S, T \subseteq \Lambda$ with $|S| = |T| = 2$.

It is enough to verify $\mu(S) = \mu(T)$ under an additional assumption that $|S \cap T| = 1$. This is because in case $S \cap T = \emptyset$ one can choose $s \in S$, $t \in T$, put $R = \{s, t\}$ and have $R \subseteq \Lambda$ while $|S \cap R| = 1 = |R \cap T|$. Then $\mu(S) = \mu(R)$ and $\mu(T) = \mu(T)$ and the transitivity implies $\mu(S) = \mu(T)$. Thus, without loss of generality assume $S = \{a, c\}$ and $T = \{b, c\}$ and construct an arbitrary tree $J$ over $\Lambda \setminus \{a, b\}$ in which $c$ is a leaf (that is, it has at most one neighbour in the tree $J$). Then the corresponding construction of two graphs will be as follows:

- the graph $G$ will have cliques $\{a, b\}$, $\{a, c\}$, all two-element cliques of $J$, and the singletons in $\Gamma$,
- the graph $H$ will have cliques $\{a, b\}$, $\{b, c\}$, all two-element cliques of $J$, and the singletons in $\Gamma$.

Since $G$ and $H$ are forests over $N$, they are chordal graphs having cliques in $\mathcal{K}$. Both graphs also have exactly $|\Lambda| - 1$ two-element cliques; these cliques $C$ are subsets of $\Lambda$ and one has $\lambda(C) = -1$ for them. The separators in the graph $G$ are either empty or singletons, which sets are outside the set $\Gamma$ of non-empty non-singletons. Thus, by (D.4), the components $m_G(T)$ for $T \in \mathcal{K} \cap \Gamma$ are either 1 or 0. Specifically, $m_G(T) = 1$ if $T$ is a (two-element) clique of $G$, which allows one to observe that the RHS of (D.3) for $m_G$ is $(-1) \cdot (|\Lambda| - 1) = 1 - |\Lambda| = 1 - |\mathcal{L} \setminus \Gamma| = \lambda_0$. The same consideration holds for $m_H$, which means that both $G$ and $H$ are tight for $\mathcal{L}$. Hence, we can include $G$ and $H$ in $\mathcal{G}$. Because, in $\mathbb{R}^{\mathcal{K} \cap \Gamma}$, one has $m_G - m_H = \delta_{\{a, c\}} - \delta_{\{b, c\}}$, it follows from (D.5) that

$$
0 = \mu_0 - \mu_0 = \sum_{T \in \mathcal{K} \cap \Gamma} \mu(T) \cdot m_G(T) - \sum_{T \in \mathcal{K} \cap \Gamma} \mu(T) \cdot m_H(T) = \sum_{T \in \mathcal{K} \cap \Gamma} \mu(T) \cdot [m_G(T) - m_H(T)] = \mu(\{a, c\}) - \mu(\{b, c\}) = \mu(S) - \mu(T),
$$

which was the goal.

A.2. Denote by $\mu^*$ the shared value $\mu(S)$ for $S \subseteq \Lambda$, $|S| = 2$.

A.3. Verify that, for every $S \in S$, $|S| \geq 3$, one has $\mu(S) = (|S| - 1) \cdot \mu^*$.

To this end, choose a node $c \in S$ and a tree $J$ over $S$ in which $c$ is a leaf. In case $\Lambda \setminus S \neq \emptyset$ also choose a node $d \in \Lambda \setminus S$ and a tree $J$ over $\Lambda \setminus S$ in which $d$ is a leaf. Then construct
• the graph $G$ which has as cliques $S$, the singletons in $\Gamma$ and, optionally in case $\Lambda \setminus S \neq \emptyset$, also $\{c, d\}$ and two-element cliques of $I$,

• the graph $H$ which has as cliques of those of $J$, the singletons in $\Gamma$ and, in case $\Lambda \setminus S \neq \emptyset$, also $\{c, d\}$ and the two-element cliques of $I$.

It is easy to observe that $G$ and $H$ are chordal graphs over $N$, and, since $S \subseteq \mathcal{K}$, their cliques are in $\mathcal{K}$. Because $H$ is a forest, the RHS in (D.3) for $m_H$ is $\mu_0$ for the same reason as mentioned in the case A.1. As concerns $G$, in $\mathbb{R}^{\mathcal{K} \cap \Upsilon}$, one has $m_G - m_H = \delta_S - \sum_{\{u,v\} \in \mathcal{J}} \delta_{\{u,v\}}$, where $\mathcal{J}$ is the set of cliques of $J$. Thus, because $\lambda(S) = 1 - |S| = \sum_{\{u,v\} \in \mathcal{J}} \lambda(\{u,v\})$, the RHS in (D.3) for $m_G$ is also $\mu_0$. Therefore, we can include both $G$ and $H$ into $\mathcal{G}$. It follows from (D.5) by subtracting the respective equations that

$$0 = \mu(S) - \sum_{\{u,v\} \in \mathcal{J}} \mu(\{u,v\}) = \mu(S) - (|S| - 1) \cdot \mu^*,$$

using the convention A.2.

A.4. Summary: we have constructed and put in $\mathcal{G}$ such graphs that (D.5) implies that there exists $\mu^*$ such that $\mu(S) = (|S| - 1) \cdot \mu^*$ for any $S \in \mathcal{S}$.

B. If $S \in \mathcal{K} \cap \Upsilon$ with $S \cap \Lambda \neq \emptyset \setminus \Gamma$ then, by (D.3), $\lambda(S) = 1 - |S \cap \Lambda| = \lambda(S \cap \Lambda)$, where we accept the convention that $\lambda(L) = 0$ whenever $L \subseteq \Lambda$, $|L| = 1$. Verify $\mu(S) = \mu(S \cap \Lambda)$ under an analogous convention $\mu(L) = 0$ for $L \subseteq \Lambda$, $|L| = 1$.

To this end, provided $\Lambda \setminus S \neq \emptyset$, choose $c \in S \cap \Lambda, d \in \Lambda \setminus S$ and a tree $J$ over $\Lambda \setminus S$ in which $d$ is a leaf. Then construct

• the graph $G$ which has cliques $S$, singletons in $\Gamma \setminus S$ and, optionally in case $\Lambda \setminus S \neq \emptyset$, also $\{c, d\}$ and two-element cliques of $I$,

• the graph $H$ whose complete sets are determined as subsets of $S \cap \Lambda$, of singletons in $\Gamma$ and, optionally in case $\Lambda \setminus S \neq \emptyset$, also of $\{c, d\}$ and two-element cliques of $J$.

Since $S \in \mathcal{K}$, one also has $S \cap \Lambda \in \mathcal{K}$; thus, the cliques of $G$ and $H$ are in $\mathcal{K}$. By (D.4), the formula for $m_G$ in $\mathbb{R}^{\mathcal{K} \cap \Upsilon}$ consists of $\delta_S$ plus an optional term

$$\sum_{\{u,v\} \in \mathcal{J}} \delta_{\{u,v\}}, \quad \text{where } \mathcal{J} \text{ is the collection of cliques of } J.$$

The formula for $m_H$ consists of $\delta_{S \cap \Lambda}$ in $\mathbb{R}^{\mathcal{K} \cap \Upsilon}$ (meaning that $\delta_{S \cap \Lambda} = 0$ in case $|S \cap \Lambda| = 1$) plus the same optional term. Hence, the RHS in (D.3) for both $m_G$ and $m_H$ is $(1 - |S \cap \Lambda|) - (|\Lambda \setminus S| = 1 - |\Lambda| = \lambda_0$ and (D.3) holds with equality for them. Therefore, we can include $G$ and $H$ into $\mathcal{G}$ and subtracting of equations (D.5) for $G$ and $H$ gives

$$0 = \sum_{T \in \mathcal{K} \cap \Upsilon} \mu(T) \cdot \left[ m_G(T) - m_H(T) \right] = \mu(S) - \mu(S \cap \Lambda),$$

where we have the convention $\mu(L) = 0$ for $L \subseteq \Lambda, |L| = 1$.

C. If $S \in \mathcal{K} \cap \Upsilon$ with $S \subseteq \Gamma$ and $S \notin \mathcal{L}^\uparrow$ then one has $\lambda(S) = 0$. Verify $\mu(S) = 0$.

To this end, choose a tree $J$ over $\Lambda$ and construct:

• the graph $G$ which has as cliques $S$, the cliques of $J$ and all the singletons in the set $\Gamma \setminus S$,

• the graph $H$ which has as cliques those of $J$ and singletons in $\Gamma$.

Both graphs are chordal and have cliques in $\mathcal{K}$. Observe, by (D.4), that, in $\mathbb{R}^{\mathcal{K} \cap \Upsilon}$, one has

$$m_H = \sum_{\{u,v\} \in \mathcal{J}} \delta_{\{u,v\}}, \quad \text{where } \mathcal{J} \text{ is the collection of two-element cliques of } J,$$
while $m_G = m_H + \delta_S$. Since $\sum_{(u,v) \in J} \lambda((u,v)) = (|A| - 1) \cdot (-1) = \lambda_0$ both graphs belong to the face determined by (D.3). Including $G$ and $H$ into $\mathcal{G}$ allows one to subtract the respective equations in (D.5) and obtain $0 = \mu(S)$.

**D.** If $S' \in \mathcal{K} \cap \Upsilon$ with $S' \subseteq \Upsilon$ and $S' \in \mathcal{L}^T$ then, by (D.3), $\lambda(S') = 1$. The details of the consideration depend on $|A|$, but in any case the next step will be needed.

**D.1.** Given $L \in \mathcal{L}$ and $S' \in \mathcal{K}$ with $L \subseteq S' \subseteq \Upsilon$ verify that $\mu(S') = \mu(L)$.

Observe that the assumption implies $L \in \Upsilon$ and we also know from (D.3) that $\lambda(S') = \lambda(L) = 1$. We choose $c \in A$ and a tree $J$ over $A$ in which $c$ is a leaf. The corresponding construction is as follows:

- the graph $G$ has cliques $S'$, $L \cup \{c\}$, all two-element cliques of $J$ and the singletons in $\Gamma \setminus S'$,
- the graph $H$ has as cliques $L \cup \{c\}$, all the two-element cliques of $J$ and the singletons in $\Gamma \setminus L$.

By the assumption $L \cup R \in \mathcal{K}$ for any $L, R \in \mathcal{L}$ we are sure that $L \cup \{c\} \in \mathcal{K}$, and, by construction, both graphs over $N$ are chordal and have cliques in $\mathcal{K}$. The formulas (D.3) for superset Möbius inversions in $\mathbb{R}^{\mathcal{K} \cap \Upsilon}$ are

$$m_H = \delta_{L \cup \{c\}} + \sum_{(u,v) \in J} \delta_{\{u,v\}}.$$  

where $J$ is the collection of two-element cliques of $J$.

$$m_G = m_H + \delta_S - \delta_L.$$  

Hence, the RHS in (D.3) for both $m_G$ and $m_H$ is

$$\lambda(L \cup \{c\}) + \sum_{(u,v) \in J} \lambda(\{u,v\}) = 0 + (-1) \cdot (|A| - 1) = \lambda_0.$$  

As $G$ and $H$ are tight for $\mathcal{L}$ they can be included into $\mathcal{G}$. By subtracting the equations (D.5) for $m_G$ and $m_H$ one gets

$$0 = \sum_{T \in \mathcal{K} \cap \Upsilon} \mu(T) \cdot [m_G(T) - m_H(T)] = \mu(S') - \mu(L).$$  

**D.2.** There exists a shared value $\mu^*$ for $\mu(S')$ for $S' \in \mathcal{K} \cap \mathcal{L}^T$ with $S' \subseteq \Upsilon$.

The crucial assumption of Lemma 5 that $|L \cup R| \leq k$ for any $L, R \in \mathcal{L}$ in this context implies that, for every pair of distinct $L, R \in \mathcal{L}$ with $L, R \subseteq \Upsilon$ one has $L \cup R \in \mathcal{K}$. This allows one to deduce $\mu(L) = \mu(L \cup R) = \mu(R)$ by the previous step D.1. Thus, there is a shared value $\mu^*$ for $\mu(L)$ for $L \in \mathcal{L}$ with $L \subseteq \Upsilon$. By applying the observation in D.1 again we obtain the desired conclusion.

**D.3.** In case $|A| \geq 2$ and $\mathcal{L} \cap \Upsilon \neq \emptyset$ observe that the shared value $\mu^*$ from the step A.2 coincides with $-\mu^c$, where $\mu^c$ is the shared value from D.2.

Because $|A| \geq 2$, we can choose different $a, b \in A$ and a tree $J$ over $\Lambda \setminus \{a\}$ in which $b$ is a leaf. Because $\mathcal{L} \cap \Upsilon \neq \emptyset$, one can also choose a set $L \in \mathcal{L}$ such that $L \subseteq \Upsilon$. The construction is as follows:

- the graph $G$ has cliques $L \cup \{a\}$, $L \cup \{b\}$, two-element cliques of $J$, and singletons in $\Gamma \setminus L$,
- the graph $H$ has as cliques $\{a,b\}$, two-element cliques of $J$ and singletons in $\Gamma$.

As $L \cup R \in \mathcal{K}$ for any $L, R \in \mathcal{L}$ we know that $L \cup \{a\}, L \cup \{b\} \in \mathcal{K}$; thus, $G$ and $H$ are chordal graphs over $N$ with cliques in $\mathcal{K}$. Moreover, by (D.4), one has in $\mathbb{R}^{\mathcal{K} \cap \Upsilon}$

$$m_H = \delta_{\{a,b\}} + \sum_{(u,v) \in J} \delta_{\{u,v\}},$$  

where $J$ is the collection of two-element cliques of $J$.

$$m_G = m_H - \delta_{\{a,b\}} + \delta_{L \cup \{a\}} + \delta_{L \cup \{b\}} - \delta_L.$$  

Hence, the RHS in (D.3) for $m_H$ is

$$\lambda(\{a,b\}) + \sum_{(u,v) \in J} \lambda(\{u,v\}) = (-1) \cdot (|A| - 1) = 1 - |A| = \lambda_0,$$

23
and, because \(-\lambda((a,b)) + \lambda(L \cup \{a\}) + \lambda(L \cup \{b\}) - \lambda(L) = +1 + 0 - 0 = 0\), the same holds for \(m_G\). Hence, \(G\) and \(H\) can be included into \(\mathcal{G}\). By subtracting the equations (D.5) for \(m_G\) and \(m_H\) one gets
\[
0 = \sum_{T \in \mathcal{K} \cap \mathcal{Y}} \mu(T) \cdot |m_G(T) - m_H(T)| = -\mu((a,b)) + \mu(L \cup \{a\}) + \mu(L \cup \{b\}) - \mu(L) = -\mu^* + 0 + 0 - \mu^* ,
\]
by the cases A.2, B and D.2. This means \(\mu^* = -\mu^*\), which was the goal.

E. Observe that if \(|\mathcal{A}| \geq 2\) then \(\mu_0 = (|\mathcal{A}| - 1) \cdot \mu^*\) and if \(|\mathcal{A}| = 1\) then \(\mu_0 = 0\).

To this end we choose a tree \(J\) over \(\Lambda\) and construct

- a graph \(G\) which has as cliques all the cliques of \(J\) and singletons in \(\Gamma\).

This is a chordal graph over \(N\) with cliques in \(\mathcal{K}\). Moreover, by (D.4), one has in \(\mathbb{R}^{\mathcal{K} \cap \mathcal{Y}}\)
\[
m_G = \sum_{|\mathcal{U}| \leq 2} \delta_{|\mathcal{U}|}, \quad \text{where } \mathcal{J} \text{ is the collection of two-element cliques of } J.
\]

Hence, the RHS in (D.3) for \(m_G\) is \(1 - |\mathcal{A}| = \lambda_0\) and \(G\) can be included into \(\mathcal{G}\). The equation (D.5) for \(m_G\) says
\[
\mu_0 = \sum_{T \in \mathcal{K} \cap \mathcal{Y}} \mu(T) \cdot m_G(T) = \sum_{|\mathcal{U}| \leq 2} \mu(|\mathcal{U}|),
\]
which is either zero, in case \(|\mathcal{A}| = 1\), or \((|\mathcal{A}| - 1) \cdot \mu^*\), in case \(|\mathcal{A}| \geq 2\) by A.2.

Now, putting the observations A-E together implies that the collection of real numbers \(\mu_0\) and \(\mu(T), T \in \mathcal{K} \cap \mathcal{Y}\), is a multiple of \(\lambda_0\) and \(\lambda(T), T \in \mathcal{K} \cap \mathcal{Y}\), which was desired. Specifically, the multiplicative factor is \(-\mu^*\) from A.2 in case \(|\mathcal{A}| \geq 2\), respectively \(\mu^*\) from D.2 in case \(|\mathcal{A}| = 1\). \(\square\)

**Appendix E. Proof of Lemma 6**

This is the result we are going to prove.

**Rephrasing Lemma 6**. Let \(M\) be a non-empty finite set. Recall that a filter is a set system \(\mathcal{F} \subseteq \mathcal{P}(M)\) closed under supersets: \(S \in \mathcal{F}, S \subseteq T \subseteq M\) implies \(T \in \mathcal{F}\). The indicator vector of a such set system \(\mathcal{F}\) will be denoted as follows:
\[
\sigma_\mathcal{F}(T) := \delta(T \in \mathcal{F}) \quad \text{for } T \subseteq M.
\]

Then the filter polytope defined by
\[
\mathcal{R} := \text{conv}\{( \sigma_\mathcal{F} \in \mathbb{R}^\mathcal{P}(M) : \mathcal{F} \subseteq \mathcal{P}(M) \text{ is a filter with } \emptyset \notin \mathcal{F}, M \in \mathcal{F} \}) \}
\]
is characterized by the following linear constraints:
\[
\sigma(\emptyset) = 0, \quad \sigma(M) = 1, \quad \sigma(B) \leq \sigma(B \cup \{a\}) \quad \text{for } a \in M, B \subseteq M \setminus \{a\}.
\]

**Proof.** The validity of (19) for \(\sigma_\mathcal{F} \in \mathcal{R}\) follows immediately from the definition of a filter. We are going to verify that every vector \(\sigma \in \mathbb{R}^\mathcal{P}(M)\) satisfying (19) is a convex linear combination of vertices of \(\mathcal{R}\). This can be shown by induction on \(s := ||T \subseteq M : \sigma(T) \neq 0||\). Observe that the inequalities (19) imply that \(0 \leq \sigma(T) \leq 1\) for any \(T \subseteq M\).

The induction premise is immediate: if \(s = 1\) then \(\sigma = \sigma_\mathcal{F}\), where \(\mathcal{F}^+ = \{M\}\) is the filter consisting of the set \(M\) only.

To verify the induction step in case \(s > 1\) we put
\[
\mathcal{F} := \{ T \subseteq M : \sigma(T) > 0 \} \quad \text{and } \beta := \min \{ \sigma(T) : T \in \mathcal{F} \} > 0
\]
and observe that \(\mathcal{F} \subseteq \mathcal{P}(M)\) is a filter with \(\emptyset \notin \mathcal{F}\) and \(M \in \mathcal{F}\). Realize that in case \(\beta = 1\) necessarily \(\sigma = \sigma_\mathcal{F}\) and the induction step is verified. Thus, assume \(\beta < 1\) in which case put
\[
\sigma' := \frac{1}{1 - \beta} \cdot [\sigma - \beta \cdot \sigma_\mathcal{F}] \in \mathbb{R}^\mathcal{P}(M) \quad \text{and have } \sigma = (1 - \beta) \cdot \sigma' + \beta \cdot \sigma_\mathcal{F}.
\]
Observe that since \( \sigma \) satisfies the constraints from (19) so does \( \sigma' \); the equations \( \sigma'(\emptyset) = 0 \) and \( \sigma'(M) = 1 \) are easy. For fixed \( a \in M \) and \( B \subseteq M \setminus \{a\} \), write
\[
(1 - \beta) \cdot [\sigma'(B \cup \{a\}) - \sigma'(B)] = \sigma(B \cup \{a\}) - \beta \cdot \sigma_F(B \cup \{a\}) - \sigma(B) + \beta \cdot \sigma_F(B)
\]
\[
\begin{align*}
&= \left\{ \begin{array}{ll}
\sigma(B \cup \{a\}) - \sigma(B) \geq 0 & \text{if } B \in \mathcal{F} \text{ or } B \cup \{a\} \notin \mathcal{F}, \\
\sigma(B \cup \{a\}) - \beta \geq 0 & \text{if } B \notin \mathcal{F} \text{ and } B \cup \{a\} \in \mathcal{F},
\end{array} \right.
\end{align*}
\]
because of the definition of \( \beta \). Now realize that \( \sigma'(T) = 0 \) for \( T \subseteq M, T \notin \mathcal{F} \), and there exists at least one \( T \in \mathcal{F} \) with \( \sigma'(T) = \beta \) and, therefore, \( \sigma'(T) = 0 \). Thus, \( s' = ||S \subseteq M : \sigma'(S) \neq 0|| < s \) and the induction hypothesis says that \( \sigma' \) is a convex combination of vertices of \( \mathcal{R} \). The formula \( \sigma = (1 - \beta) \cdot \sigma' + \beta \cdot \sigma_F \) then completes the proof of the induction step.

References