International Journal of Approximate Reasoning 138 (2021) 188–203

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

The dual polyhedron to the chordal graph polytope and the rebuttal of the chordal graph conjecture ☆

Milan Studený [a,*], James Cussens [b], Václav Kratochvíl [a]

[a] Czech Academy of Sciences, The Institute of Information Theory and Automation, Prague, Pod Vodárenskou věží 4, 18200, Czech Republic
[b] Department of Computer Science, University of Bristol, Merchant Venturers Building, Woodland Road, Bristol, BS8 1UB, UK

ARTICLE INFO

Article history:
Received 14 January 2021
Received in revised form 21 May 2021
Accepted 26 July 2021
Available online 26 August 2021

Keywords:
Learning decomposable models
Chordal graph polytope
Clutter inequalities
Dual polyhedron

ABSTRACT

The integer linear programming approach to structural learning of decomposable graphical models led us earlier to the concept of a *chordal graph polytope*. An open mathematical question motivated by this research is what is the minimal set of linear inequalities defining this polytope, i.e. what are its facet-defining inequalities, and we came up in 2016 with a specific conjecture that it is the collection of so-called *clutter* inequalities. In this theoretical paper we give an implicit characterization of the minimal set of inequalities. Specifically, we introduce a *dual polyhedron* (to the chordal graph polytope) defined by trivial equality constraints, simple monotonicity inequalities and certain inequalities assigned to incomplete chordal graphs. Our main result is that the vertices of this polyhedron yield the facet-defining inequalities for the chordal graph polytope. We also show that the original conjecture is equivalent to the condition that all vertices of the dual polyhedron are zero-one vectors. Using that result we disprove the original conjecture: we find a vector in the dual polyhedron which is not in the convex hull of zero-one vectors from the dual polyhedron.

1. Introduction

The source of motivation for this paper is learning *decomposable graphical models*, which are assigned to chordal undirected graphs. These are fundamental graphical statistical models [1] and one of the reasons for this is that their elegant mathematical properties are at the core of the well-known method of local computation [2]. Decomposable models can also be interpreted as special cases of Bayesian network models [3] assigned to directed acyclic graphs.

In particular, most of the methods for structural learning of decomposable models follow the standard methods for learning Bayesian networks [4]. We are specifically interested in the *integer linear programming* (ILP) approach to structural learning of decomposable models. This approach is a special case of the score-based approach: methods of ILP are used with the goal to maximize some additively decomposable score, like the BIC score [5] or the BDeu score [6]. The idea behind the ILP approach is to encode graphical models by vectors whose components are integers in such a way that the standard scores turn into linear functions of the vector representatives, up to a constant term. Note that the number of components of such vectors is inevitably exponential in the number of nodes of graphs. There are also different ways to encode Bayesian network models by vectors and these ways were compared in [7].

☆ This paper belongs to the virtual special issue on the 10th International Conference on Probabilistic Graphical Models.

* Corresponding author.
E-mail addresses: studeny@utia.cas.cz (M. Studený), james.cussens@bristol.ac.uk (J. Cussens), velorex@utia.cas.cz (V. Kratochvíl).

https://doi.org/10.1016/j.ijar.2021.07.014
0888-613X/

The specific ILP approach to learning of decomposable models discussed in this paper is based on encoding the models by vectors named *characteristic imsets*; these vectors were earlier introduced and tested in the context of learning Bayesian networks [8,9]. This mode of representation leads to a particularly elegant way of encoding chordal graphs. Nonetheless, an additional invertible linear mapping, named *superset Möbius inversion* in this paper, allows one to transform characteristic imsets into other vector representatives which have only a few non-zero components; these components correspond to cliques and separators of the graphs and are, therefore, close to junction tree representations of the graphs considered by other authors (see Section 1.2 below).

The so-called *chordal graph polytope* was formally defined by two of us in a 2016 conference paper [10] as the convex hull of all characteristic imsets for chordal graphs over a fixed set N of nodes, $|N| \geq 2$. This polytope had been introduced even earlier under a longer name by Lindner [11] in her 2012 thesis, where she derived some basic observations on it.

We have also introduced in [10] special linear *clutter inequalities* valid for the vectors in the polytope, which correspond to singleton-containing clutters (= classes of inclusion incomparable sets, alternatively named Sperner families or anti-chains) of subsets of N. Moreover, it was shown in our later 2017 journal paper [12] that each clutter inequality is facet-defining for the chordal graph polytope; that is, it belongs to the minimal set of linear inequalities defining the polytope. Note in this context that we do not distinguish between an inequality and its multiple by a positive factor and that the uniqueness of the minimal defining set of linear inequalities is relative to the affine (= shifted linear) space generated by the polytope.

Let us explain here the motivation for characterizing the minimal set of inequalities delimiting the chordal graph polytope. Assuming that one has such a theoretical characterization of this polytope at one's disposal one would be able to re-formulate common statistical learning tasks for decomposable models in the form of *linear programming* (LP) problems instead of in the form of ILP problems, as is the present state of art. The main point here is that LP problems are polynomially solvable whereas ILP problems are NP-hard. Thus, while a number of highly efficient methods are available to solve LP problems, the methods for solving ILP problems are based on repeated solving of LP problems and may be computationally impractical. A theoretical characterization of facets of a domain polytope can also be utilized to design an efficient implementation of the *cutting plane method*, which is one of standard methods to solve ILP problems.

Following the observations from [10,12] we have raised a natural *chordal graph conjecture* there, saying that the minimal set of inequalities defining the chordal graph polytope consists of the clutter inequalities and one additional *lower bound inequality* which requires the non-negativity of the component for N. The conjecture has been confirmed computationally in case $|N| \leq 5$ and since then we have been trying quite intensively either to confirm or disprove it.

Note that if the conjecture had been confirmed then it would have established a kind of duality relationship between two important combinatorial categories, namely between chordal undirected graphs over N and clutters of subsets of N containing singletons. Therefore, the raised question was also of fundamental theoretical importance from the point of view of combinatorics. The conjecture was, however, refuted in the case $|N| = 6$ in our recent conference paper [13] and this journal paper is an extended version of that conference contribution containing complete proofs of all stated results.

1.1. Highlights in this paper

Specifically, below we introduce a certain bounded polyhedron in $[0,1]^{\mathcal{P}(N)}$, where $\mathcal{P}(N)$ is the power set of N, specified by a few elementary monotonicity inequalities and certain special *inequalities assigned to incomplete chordal graphs* over N. The main result of the paper is that the vertices of this *dual polyhedron* yield all facet-defining inequalities for the chordal graph polytope; thus, they give rise to the minimal set of inequalities characterizing it.

A further observation is that the original chordal graph conjecture is equivalent to the condition that all vertices of the dual polyhedron are zero-one vectors. We also show that the zero-one vertices of the dual polyhedron correspond to singleton-containing clutters of subsets of N. In fact, each of these vertices is the indicator of supersets of sets in such a clutter; that is, it is the indicator of the *set filter* generated by such a clutter. The latter observation allows us to disprove the original chordal graph conjecture: we find a vector in the dual polyhedron for $|N| = 6$ which is not in the convex hull of the indicators of above-mentioned set filters.

Thus, an optimistic conjecture that facets of the chordal graph polytope over N correspond to special systems of subsets of N was refuted. On the other hand, the findings from this paper do not ruin completely the hope in characterizing facets of the chordal graph polytope and their use in the ILP approach to learning decomposable models – see the discussion in Conclusions.

1.2. Other approaches to learning decomposable models

Our approach is not the only ILP approach to learning of decomposable models. Other authors have used vector representatives of these models different from ours. Sesh Kumar and Bach [14] employed special codes for junction trees that correspond to chordal graphs, while Pérez, Blum and Lozano [15,16] used zero-one vector encodings of certain special coarsenings of maximal hypertrees. Note in this context that they all have been interested in learning decomposable graphical models with a given clique size limit.

There are also approaches to structural learning of decomposable models which are not based on ILP but some of them use encodings of junction trees as well. Corander et al. [17] employed a constraint satisfaction approach and expressed their search space of models in terms of logical constraints. Kangas, Niinimäki and Koivisto [18,19] applied the idea of decomposing junction trees into subtrees and used methods of dynamic programming. Rantanen, Hyttinen, and Järvisalo [20] have then employed a branch and bound method and integrated dynamic programming as well.

Let us note in this context that the above mentioned superset Möbius inversions of characteristic imsets are naturally related to the junction tree representatives of decomposable models used both by Sesh Kumar and Bach [14] and by Kangas, Niinimäki and Koivisto [18,19]. They are also linearly related to the so-called *standard imsets* for chordal graphs from [21, § 7.2.2]. We emphasize those links here because the coefficients of the inequalities assigned to chordal graphs in our dual formulation of the problem are just components of these vector representatives (up to a multiplicative factor).

1.3. Structure of this paper

In this paper we omit the concepts related to statistical learning of graphical statistical models because these concepts are not needed to present our result; the reader can find them in [12]. We assume that the reader is familiar with elementary concepts from polyhedral geometry which can be found in standard textbooks like [22], [23] or [24].

In Section 2 we recall basic definitions and facts. In Section 3 we introduce the concept of a dual polyhedron and formulate our theoretical results; their proofs are, however, moved to Appendices. Those results allow us to disprove the chordal graph conjecture, which is done by means of Example 7 in Section 4. In Conclusions (Section 5) we summarize our findings and discuss potential future research directions.

Let us recall here what is the addition of the present paper compared to the original conference paper [13]. Because of a page limit for [13] the proofs of some crucial results were skipped there or replaced by sketchy hints only. In this paper detailed proofs are given; particularly concerning substantial Example 7 and (newly proved) Theorem 2. To make the paper better to follow we have also incorporated new examples in Section 2.3 illustrating the role of clutter inequalities. The open questions from Conclusions are also new.

2. Basic concepts

Let N be a finite set of *variables* which correspond to the nodes of our graphs and in the statistical context are interpreted as random variables. To avoid the trivial case we assume $n := |N| \geq 2$. Let $\mathcal{P}(N) := \{S: \; S \subseteq N\}$ denote the power set of N. We call a subset $S \subseteq N$ a *singleton* if $|S| = 1$. Given a set system $\mathcal{L} \subseteq \mathcal{P}(N)$, the symbol $\bigcup \mathcal{L}$ will denote the union of sets from $\mathcal{L}$. Given a predicate $\mathbf{P}$, the symbol $\delta(\mathbf{P})$ will denote the zero-one indicator of $\mathbf{P}$, that is, $\delta(\mathbf{P}) = 1$ if $\mathbf{P}$ holds and $\delta(\mathbf{P}) = 0$ if $\mathbf{P}$ does not hold.

2.1. Chordal graphs

An undirected graph G *over* N (= a graph G having N as the set of nodes) is called *chordal* if every cycle in G 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 in G if every two distinct nodes from S are connected by an edge in G. The maximal complete sets with respect to inclusion are called the *cliques* of G. A (chordal) graph G over N is complete if N is a clique, otherwise it is called *incomplete*.

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):

$$\forall\, i \geq 2 \;\; \exists\, j < i \quad \text{such that } \; S_i := C_i \cap \left(\bigcup_{\ell < i} C_\ell \right) \subseteq C_j \,.$$

The sets $S_i = C_i \cap (\bigcup_{\ell<i} C_\ell)$, $i = 2, \ldots, m$, are the respective separators. The *multiplicity* $\nu_G(S)$ 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 [21, Lemma 7.2]. In the sequel, the collection of *cliques of* G will be denoted by $\mathcal{C}(G)$ and the collection of its *separators* by $\mathcal{S}(G)$.

A *junction tree* for G is a hypertree $\mathcal{J}$ having $\mathcal{C}(G)$ as the set of hypernodes in $\mathcal{J}$ and satisfying the condition that, for every pair $C, K \in \mathcal{C}(G)$, the intersection $C \cap K$ is contained in every clique on the (unique) path between C and K in $\mathcal{J}$. Another equivalent definition of a chordal graph is that it is an undirected graph which has a junction tree; see [2, Theorem 4.6]. A hyperedge between C and K in a junction tree $\mathcal{J}$ can be labeled by the intersection $C \cap K$ and another well-known fact is that the labels of these hyperedges are just the separators of G and every separator S occurs as many times in the junction tree $\mathcal{J}$ as its multiplicity $\nu_G(S)$.

2.2. Characteristic imsets for chordal graphs and the chordal graph polytope

Given a chordal graph G over N, the *characteristic imset* of G is a zero-one vector c_G whose components are indexed by subsets S of N:

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

Thus, c_G is formally a vector in $\mathbb{R}^{\mathcal{P}(N)}$ whose components for the empty set and singletons have always the value 1. Nonetheless, the roles of the empty set and singletons in our later linear inequalities for characteristic imsets differ. While the component for the empty set plays no role in our inequalities, it is useful to distinguish between components for different singletons because this step allows us to identify our inequalities with certain set systems; see later Example 2 to illustrate that. Therefore, we will understand every characteristic imset as a vector c in the linear space $\mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$ belonging to its affine subspace

$$\mathsf{A} := \{\mathsf{c} \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}} \,:\, \mathsf{c}(\{i\}) = 1 \quad \text{for all } i \in N\}$$

specified by those equality constraints. Let us introduce the *chordal graph polytope* D_N over N as follows:

$$D_N \;:=\; \mathrm{conv}\,(\,\{\mathsf{c}_G \,:\, G \text{ chordal graph over } N\}\,)$$

where $\mathrm{conv}\,(\cdot)$ denotes the convex hull. Since A is the affine hull of D_N the dimension of D_N is $2^n - n - 1$, where $n = |N|$. This is because for any $A \subseteq N$ with $|A| \geq 2$ one has $\mathsf{c}_G \in D_N$ for the (chordal) graph G having cliques A and $\{i\}$ for $i \in N \setminus A$; see also [12, § 2.5].

2.3. Clutter inequalities and the chordal graph conjecture

By a *clutter* we mean any set system $\mathcal{L} \subseteq \mathcal{P}(N)$ such that the sets in $\mathcal{L}$ are inclusion incomparable:

$$\text{if } L, R \in \mathcal{L} \quad \text{then} \quad L \subseteq R \text{ implies } L = R.$$

Note that the union $\bigcup \mathcal{L}$ of sets in a clutter need not be the whole set N. A (set) *filter* is a set system $\mathcal{F} \subseteq \mathcal{P}(N)$ closed under supersets:

$$\text{if } S \in \mathcal{F} \text{ and } S \subseteq T \subseteq N \quad \text{then} \quad T \in \mathcal{F}.$$

Clutters and filters are in mutual correspondence: any clutter $\mathcal{L}$ generates a filter $\mathcal{L}^{\uparrow} := \{T \subseteq N \,:\, \exists\, L \in \mathcal{L} \quad L \subseteq T\}$ and conversely, given a filter $\mathcal{F}$, the class $\mathcal{F}_{\min}$ of inclusion minimal sets in $\mathcal{F}$ is a clutter generating $\mathcal{F}$. Note that a clutter of non-empty sets contains a singleton iff the corresponding filter contains a singleton.

Notational convention. In our examples, clutters of subsets of N will be denoted by the lists of sets in the clutters separated by straight lines. The sets are encoded by lists of their elements without commas. Thus, for example, a clutter $\mathcal{L} = \{\ \{a,b\},\ \ \{a,c\},\ \ \{d\}\ \}$ will be denoted by $|d|ab|ac|$. With a little abuse of notation, we will also use the symbol for a clutter to identify the corresponding filter. Thus, $\lambda_{|d|ab|ac|} \in \mathbb{R}^{\mathcal{P}(N)}$ will denote the indicator of the filter generated by the clutter $|d|ab|ac|$. Chordal graphs over N will be analogously denoted by (all) their cliques separated by colons; thus, the empty graph over $N = \{a,b,c,d\}$ will be denoted by $a:b:c:d$ and the graph $a - b - c - d$ will be denoted by $ab:bc:cd$.

Definition 1 *(clutter inequality).* Given a clutter $\mathcal{L} \subseteq \mathcal{P}(N)$ which contains at least one singleton the corresponding *clutter inequality* for $\mathsf{c} \in \mathsf{A} \subseteq \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$ has the form

$$1 \leq \sum_{\emptyset \neq \mathcal{B} \subseteq \mathcal{L}} (-1)^{|\mathcal{B}|+1} \cdot \mathsf{c}\left(\bigcup \mathcal{B}\right). \tag{1}$$

Recall that because we assume $\mathsf{c} \in \mathsf{A}$ one has to substitute $\mathsf{c}(L) = 1$ in (1) whenever $L \subseteq N$ is a singleton. Note also that the inequality for a clutter consisting of only one (singleton) set is superfluous: then $\mathcal{L} = \{L\}$ with $L \subseteq N$, $|L| = 1$, and (1) holds with equality since $\mathsf{c} \in \mathsf{A}$.

One can re-write (1) in a standardized form:

$$1 \;\leq\; \sum_{\emptyset \neq S \subseteq N} \kappa_{\mathcal{L}}(S) \cdot \mathsf{c}(S) \qquad \text{where } \kappa_{\mathcal{L}}(S) = \sum_{\emptyset \neq \mathcal{B} \subseteq \mathcal{L} \,:\, \bigcup \mathcal{B} = S} (-1)^{|\mathcal{B}|+1} \text{ for any } S \subseteq N. \tag{2}$$

This rewriting implies that the coefficients $\kappa_{\mathcal{L}}(-)$ have to vanish outside the class $\mathcal{U}(\mathcal{L}) := \{\bigcup \mathcal{B} \,:\, \emptyset \neq \mathcal{B} \subseteq \mathcal{L}\}$ of unions of sets from $\mathcal{L}$. As shown in [12, Lemma 1], they can be computed recursively within this class as follows:

$$\kappa_{\mathcal{L}}(S) = 1 - \sum_{T \in \mathcal{U}(\mathcal{L}) \colon T \subset S} \kappa_{\mathcal{L}}(T) \qquad \text{for any } S \in \mathcal{U}(\mathcal{L}).$$

Hence, $\kappa_{\mathcal{L}}(L) = 1$ for $L \in \mathcal{L}$. The example below illustrates the procedure.

Example 1. Take $N = \{a,b,c,d\}$ and a clutter $\mathcal{L}$ specified by $|d|ab|ac|$. The fact that $\kappa_{\mathcal{L}}(L) = 1$ for $L \in \mathcal{L}$ gives $\kappa_{\mathcal{L}}(\{a,b\}) = \kappa_{\mathcal{L}}(\{a,c\}) = \kappa_{\mathcal{L}}(\{d\}) = 1$. The remaining elements in $\mathcal{U}(\mathcal{L})$ are $\{a,b,c\}$, $\{a,b,d\}$, $\{a,c,d\}$, and N. The recursive formula above applied to $\{a,b,c\}$ yields

$$\kappa_{\mathcal{L}}(\{a,b,c\}) \;=\; 1 - \kappa_{\mathcal{L}}(\{a,b\}) - \kappa_{\mathcal{L}}(\{a,c\}) = 1 - 1 - 1 = -1\,.$$

Analogously, $\kappa_{\mathcal{L}}(\{a,b,d\}) = \kappa_{\mathcal{L}}(\{a,c,d\}) = -1$. Finally, N has all other sets in $\mathcal{U}(\mathcal{L})$ as proper subsets which gives

$$\kappa_{\mathcal{L}}(N) \;=\; 1 - \kappa_{\mathcal{L}}(\{a,b,c\}) - \kappa_{\mathcal{L}}(\{a,b,d\}) - \kappa_{\mathcal{L}}(\{a,c,d\}) - \kappa_{\mathcal{L}}(\{a,b\}) - \kappa_{\mathcal{L}}(\{a,c\}) - \kappa_{\mathcal{L}}(\{d\}) = 1 + 3 - 3 = 1\,.$$

Taking into consideration that $\mathsf{c}(\{d\}) = 1$ one gets from (2):

$$1 \le \mathsf{c}(\{a,b\}) + \mathsf{c}(\{a,c\}) + 1 - \mathsf{c}(\{a,b,c\}) - \mathsf{c}(\{a,b,d\}) - \mathsf{c}(\{a,c,d\}) + \mathsf{c}(N)\,,$$

which can be re-written in the form

$$\mathsf{c}(\{a,b,c\}) + \mathsf{c}(\{a,b,d\}) + \mathsf{c}(\{a,c,d\}) \le \mathsf{c}(\{a,b\}) + \mathsf{c}(\{a,c\}) + \mathsf{c}(N)\,.$$

The next example is to illustrate that singletons are needed to distinguish between different (clutter) inequalities.

Example 2. Assume $N = \{a,b,c,d\}$. Then both the inequality $\mathsf{c}(abc) \le \mathsf{c}(ab)$ and the inequality $\mathsf{c}(abd) \le \mathsf{c}(ab)$ are facet-defining for the chordal graph polytope D_N. Although these two different inequalities do not involve singletons, to distinguish between them we associate them, by the procedure from Example 1, with different clutters $|c|ab|$ and $|d|ab|$ which *do involve the singletons*. We hope that this example shows the reader why it is useful to distinguish between components of characteristic imsets c for different singletons even though we have a natural convention that $\mathsf{c}(\{i\}) = 1$ for $i \in N$.

The next two examples are to illustrate how the clutter inequalities work.

Example 3. Take $N = \{a,b,c,d\}$ and a clutter $\mathcal{L}$ specified by $|a|b|c|d|$. The procedure from Example 1 leads to the inequality

$$\begin{aligned}&\mathsf{c}(\{a,b\}) + \mathsf{c}(\{a,c\}) + \mathsf{c}(\{a,d\}) + \mathsf{c}(\{b,c\}) + \mathsf{c}(\{b,d\}) + \mathsf{c}(\{c,d\}) + \mathsf{c}(\{a,b,c,d\})\\ &\qquad\le 3 + \mathsf{c}(\{a,b,c\}) + \mathsf{c}(\{a,b,d\}) + \mathsf{c}(\{a,c,d\}) + \mathsf{c}(\{b,c,d\})\,.\end{aligned}$$

In the case of a chordal graph G determined by $ad:abc$ the substitution of the respective characteristic imset c_G gives $4 \le 3+1$, which means the inequality holds with equality. On the other hand, if one considers a *non-chordal* graph H over N "determined" by $ab:bc:cd:ad$, that is, a graph with a pure 4-cycle, then the substitution of the indicator of complete sets in H in the inequality gives $4 \le 3+0$, which is an *invalid* inequality. Thus, the clutter inequality corresponding to $|a|b|c|d|$ can possibly be interpreted as an inequality which excludes cycles of the length 4. Recall from [12, § 5.1] that this inequality was earlier known as the *cluster inequality* corresponding to a 4-element set.

One can analogously interpret the clutter inequality specified by $|a|b|c|$, that is, the cluster inequality corresponding to $\{a,b,c\}$, as an inequality forbidding a 3-cycle over $\{a,b,c\}$. This straightforward interpretation, however, makes sense only in case of zero-one vectors c. To forbid fractional vectors "hiding" cycles additional clutter inequalities are needed.

Example 4. Take $N = \{a,b,c,d\}$ and a clutter $\mathcal{L}$ specified by $|c|d|ab|$. The respective clutter inequality has the form

$$\mathsf{c}(\{c,d\}) + \mathsf{c}(\{a,b,c\}) + \mathsf{c}(\{a,b,d\}) \le 1 + \mathsf{c}(\{a,b\}) + \mathsf{c}(\{a,b,c,d\})\,.$$

In the case of a chordal graph G determined by $ad:abc$ the substitution of the respective characteristic imset c_G gives $1 \le 1+1$; thus, the inequality does not hold with equality but with the surplus 1. Let us note in this context that the surplus of the inequality given by a clutter $\mathcal{L}$ if applied to the characteristic imset of a chordal graph G has an elegant interpretation in terms of (any) junction tree $\mathcal{J}$ for G. Specifically, it is related to the number of components of a certain sub-forest of $\mathcal{J}$ determined by $\mathcal{L}$; see later formula (7) and the proof of Lemma 1 in Appendix A for details. On the other hand, if one considers a fractional vector $\mathsf{c}^* := \frac{1}{3}\cdot\mathsf{c}_G + \frac{2}{3}\cdot\mathsf{c}'$, where G is the chordal graph specified by $a:b:c:d$ and c' is the "zero-one indicator" of subsets in $cd:abc:abd$ then the substitution of c^* in the considered inequality gives $2 \le 1+\frac{2}{3}$, which is an *invalid* inequality. Note in this context that this is the only clutter inequality for N which excludes the fractional vector c^* from D_N.

Our conjecture was that all facet-defining inequalities for D_N were as follows.

Conjecture 1 *(chordal graph conjecture). For any $n = |N| \ge 2$, the least set of inequalities for $\mathsf{c} \in \mathsf{A}$ defining D_N consists of the lower bound inequality $0 \le \mathsf{c}(N)$ and the inequalities (1) for those clutters $\mathcal{L}$ of subsets of N that contain at least one singleton and at least one another set.*

Note that the conjecture was confirmed computationally for $|N| \leq 5$.

2.4. Two different Möbius inversions

The linear transformations of vectors (with components indexed by subsets of N) considered in this paper are based on a well-known combinatorial inclusion-and-exclusion principle, which is usually interpreted as a special case of (generalized) Möbius inversion formula [25, § 3.7]. Thus, we follow common combinatorial terminology and call these transformations *Möbius inversions*. Nevertheless, we have to distinguish two different transformations of this kind.

Given a vector $\mathsf{c} \in \mathbb{R}^{\mathcal{P}(N)}$, its *superset Möbius inversion* is the vector $\mathsf{m} \in \mathbb{R}^{\mathcal{P}(N)}$ given by

$$\mathsf{m}(T) := \sum_{S:\, T \subseteq S} (-1)^{|S\setminus T|} \cdot \mathsf{c}(S) \quad \text{for any } T \subseteq N. \tag{3}$$

The inverse formula to (3) is $\mathsf{c}(S) = \sum_{T:\, S\subseteq T} \mathsf{m}(T)$ for any $S \subseteq N$, which is easy to verify by a direct substitution and re-arranging sums. Note that in both formulas we sum over *supersets* of the set for which we compute the value. By [12, Lemma 3], the superset Möbius inversion of the characteristic imset c_G of a chordal graph G over N has the form

$$\mathsf{m}_G(T) = \sum_{C \in \mathcal{C}(G)} \delta(T = C) - \sum_{S \in \mathcal{S}(G)} \nu_G(S) \cdot \delta(T = S) \quad \text{for } T \subseteq N\,. \tag{4}$$

Every linear inequality for c_G can be re-written in terms of its superset Möbius inversion m_G and conversely; however, the relation of the respective coefficient vectors is given by another Möbius inversion. Specifically, given a vector $\lambda \in \mathbb{R}^{\mathcal{P}(N)}$ its *subset Möbius inversion* is given by

$$\kappa_\lambda(S) := \sum_{T:\, T \subseteq S} (-1)^{|S\setminus T|} \cdot \lambda(T) \quad \text{for any } S \subseteq N\,. \tag{5}$$

Note that the inverse formula to (5) is $\lambda(T) = \sum_{S:\, S\subseteq T} \kappa_\lambda(S)$ for any $T \subseteq N$ and that we sum over *subsets* in both these formulas. Provided that $\lambda \in \mathbb{R}^{\mathcal{P}(N)}$ is a coefficient vector for a linear inequality and the vector $\mathsf{m} \in \mathbb{R}^{\mathcal{P}(N)}$ is the superset Möbius inversion of a vector $\mathsf{c} \in \mathbb{R}^{\mathcal{P}(N)}$ one has:

$$\sum_{T \subseteq N} \lambda(T) \cdot \mathsf{m}(T) = \sum_{S \subseteq N} \kappa_\lambda(S) \cdot \mathsf{c}(S)\,, \quad \text{which particularly implies} \tag{6}$$
$$1 \leq \sum_{T \subseteq N} \lambda(T) \cdot \mathsf{m}(T) \iff 1 \leq \sum_{S \subseteq N} \kappa_\lambda(S) \cdot \mathsf{c}(S)\,.$$

The verification of this fact can be done by direct substitution into the formulas and re-arranging the sums; it is left to the reader. A special case of the equivalence in (6) is a standard re-writing of the clutter inequalities in terms of superset Möbius inversion m_G from [12, Lemma 2]. Specifically, given a clutter $\mathcal{L} \subseteq \mathcal{P}(N)$ (containing a singleton), consider the indicator $\lambda \in \mathbb{R}^{\mathcal{P}(N)}$ of the filter $\mathcal{L}^{\uparrow}$ generated by $\mathcal{L}$. Then the subset Möbius inversion κ_λ of λ is just the coefficient vector $\kappa_{\mathcal{L}}$ of the respective clutter inequality:

$$1 \leq \sum_{T \subseteq N} \delta(T \in \mathcal{L}^{\uparrow}) \cdot \mathsf{m}_G(T) \qquad \iff \qquad 1 \leq \sum_{S \subseteq N} \kappa_{\mathcal{L}}(S) \cdot \mathsf{c}_G(S)\,. \tag{7}$$

In our later dual formulation of the problem we are going to assign certain linear inequalities to incomplete chordal graphs G over N. The coefficients of those inequalities are given by slightly modified superset Möbius inversions m_G; specifically, given an *incomplete chordal graph* G over N, we introduce a vector $\overline{\mathsf{m}}_G \in \mathbb{R}^{\mathcal{P}(N)}$ as follows:

$$\overline{\mathsf{m}}_G(N) = -1 \qquad \text{and} \quad \overline{\mathsf{m}}_G(S) := \mathsf{m}_G(S) \quad \text{for remaining } S \subset N. \tag{8}$$

Note for a reader interested in earlier occurrences of concepts defined here in the literature that the vector $\overline{\mathsf{m}}_G$ given by (8) is nothing else than (-1) multiple of the *standard imset* for the chordal graph G as introduced in [21, § 7.2.2].

3. Dual formulation of the problem

We assign the following inequality for vectors $\lambda \in \mathbb{R}^{\mathcal{P}(N)}$ to any *incomplete chordal graph* G over N:

$$\sum_{C \in \mathcal{C}(G)} \lambda(C) - \sum_{S \in \mathcal{S}(G)} \nu_G(S) \cdot \lambda(S) - \lambda(N) \geq 0\,, \tag{9}$$
$$\text{that is, } \langle \overline{\mathsf{m}}_G, \lambda \rangle := \sum_{S \subseteq N} \overline{\mathsf{m}}_G(S) \cdot \lambda(S) \geq 0 \ \text{ using the formulas (4) and (8)}.$$

Definition 2 *(dual polyhedron to the chordal graph polytope).* Let us define the *dual polyhedron* $\mathsf{P} \subseteq \mathbb{R}^{\mathcal{P}(N)}$ as the set of vectors $\lambda \in \mathbb{R}^{\mathcal{P}(N)}$ satisfying

- $\lambda(\emptyset) = 0$, $\lambda(N) = 1$,
- simple monotonicity inequalities $\lambda(N) - \lambda(N \setminus \{i\}) \geq 0$ for $i \in N$, and
- the above inequalities (9) for all incomplete chordal graphs over N.

Remark This is to explain the motivation for our terminology. There is a concept of *duality* for polytopes, which is one of the standard concepts in polyhedral geometry, see [26, § 10]. More specifically, two polytopes are dual to each other if their face-lattices (= the lattices of their faces) are anti-isomorphic. Each polytope has a dual polytope in this sense [26, Theorem 10.2]; a standard such construction is based on the concept of a "polar" set [27, Corollary 2.13]. The duality condition implies that the vertices of one polytope are in one-to-one correspondence to the facets of the other polytope. And this property was our motivation: with a little abuse of terminology we call the above defined set P the *dual* polyhedron *to the chordal graph polytope* because, as one can show (see Theorem 1 below) that its vertices are in one-to-one correspondence to facets of the dominant facet of the chordal graph polytope D_N, determined by the lower bound inequality; this dominant facet of D_N is the convex hull of all (characteristic imsets of) incomplete chordal graphs over N. Because the set P is defined in terms of inequalities we call it a *polyhedron* rather than a "polytope", by which is meant the convex hull of finitely many vectors. Note that, as we show below (Lemma 1), P is a bounded polyhedron, and, therefore, a polytope, by a fundamental result in polyhedral geometry [26, Corollary 8.7].

We prove in Appendix A the following.

Lemma 1 *(basic facts on the dual polyhedron). Every $\lambda \in \mathsf{P}$ is non-decreasing relative to inclusion: $\lambda(S) \leq \lambda(T)$ whenever $S \subseteq T$. In particular, the set $\mathsf{P} \subseteq [0,1]^{\mathcal{P}(N)}$ is a non-empty bounded polyhedron. One has $\lambda \in \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}$ iff λ is the indicator of a singleton-containing filter $\mathcal{F} \subseteq \mathcal{P}(N)$ with $\emptyset \notin \mathcal{F}$. Moreover, $\{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P} \subseteq \mathrm{ext}(\mathsf{P})$, where $\mathrm{ext}(\mathsf{P})$ denotes the set of vertices (= extreme points) of* P.

Thus, indicators of non-degenerate singleton-containing filters are always vertices of the dual polyhedron. Note that these are all the vertices of P in the case $|N| \leq 5$, a fact which was confirmed computationally.

Example 5. Consider the case $N = \{a,b,c\}$. Then the dual polyhedron P has dimension $6 = 2^3 - 2$ and is specified by 10 inequalities breaking into 4 permutational types. One has 3 simple monotonicity inequalities of the form $\lambda(N) \geq \lambda(N \setminus \{i\})$ and 7 inequalities assigned to incomplete chordal graphs falling into 3 permutational types: $i:j:k$, $i:jk$, and $ij:ik$. The number of vertices of P is also 10 and they fall into 4 permutational types. All of them are the indicators of singleton-containing filters; their types are (denoted by) $|i|$, $|i|jk|$, $|i|j|$, and $|i|j|k|$.

Here is the main result of the paper, proved in Appendix B.

Theorem 1 *(dual characterization of the chordal graph polytope). Assume $|N| \geq 2$. Let $\mathrm{ext}(\mathsf{P})$ denote the set of vertices of the dual polyhedron. Given $\mathsf{c} \in \mathsf{A}$, one has $\mathsf{c} \in D_N$ if and only if $\mathsf{c}(N) \geq 0$ and the inequalities*

$$\sum_{S \subseteq N} \kappa_\lambda(S) \cdot \mathsf{c}(S) \geq 1 \tag{10}$$

hold for every $\lambda \in \mathrm{ext}(\mathsf{P})$.

Recall that one has $\kappa_\lambda(\emptyset) = 0$ for any $\lambda \in \mathsf{P}$, which means that it does not matter whether one allows the empty set component in (10) or not. In particular, (10) can equivalently be written in the form

$$\sum_{\emptyset \neq S \subseteq N} \kappa_\lambda(S) \cdot \mathsf{c}(S) \geq 1\,.$$

Note also in this context that $\mathsf{c} \in \mathsf{A}$ means $1 = \mathsf{c}(\{i\}) = \sum_{S \subseteq N} \kappa_{\lambda_{|i|}}(S) \cdot \mathsf{c}(S)$ for any $i \in N$ (see Section 2.4). Thus, because $\lambda_{|i|} \in \mathrm{ext}(\mathsf{P})$ for any $i \in N$, these $|N|$ inequalities (10) for $\lambda_{|i|}$ hold with equality. The inequalities (10) for the remaining $\lambda \in \mathrm{ext}(\mathsf{P})$, however, appear to be non-trivial (= do not hold with equality). Theorem 1 can be interpreted as an implicit characterization of facet-defining inequalities for D_N: it can be shown by extending the arguments in the proof from Appendix B that the non-trivial inequalities (10) are, in fact, facet-defining for D_N. In particular, the minimal set of linear inequalities defining D_N can be obtained on basis of $\mathrm{ext}(\mathsf{P})$.

Example 6. Consider again $N = \{a,b,c\}$. As mentioned in Example 5, P has 10 vertices. Those 3 of them of the type $|i|$ yield equality constraints and the remaining 7 ones, those of types $|i|jk|$, $|i|j|$, and $|i|j|k|$, lead to respective *clutter inequalities* (of

3 types). These are facet-defining for the chordal graph polytope D_N. By adding the lower bound inequality $\mathsf{c}(N) \geq 0$ the list of facet-defining inequalities for D_N becomes complete.

The second substantial result, proved in Appendix C, is as follows.

Theorem 2 *(dual formulation of the chordal graph conjecture). The chordal graph conjecture holds for $|N| \geq 2$ if and only if* $\{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P} = \mathrm{ext}\,(\mathsf{P})$.

4. Rebuttal of the conjecture

We found a counter-example to the validity of the chordal graph conjecture in the case $|N| = 6$. The argument is based on Theorem 2: we find a vector in the dual polyhedron P which is *not* in the convex hull of the zero-one vectors from P. Recall in this context that, by Lemma 1, these zero-one vectors from P are none other than the indicators of non-degenerate singleton-containing filters.

Example 7. Take $N = \{a,b,c,d,e,f\}$. We present an example of $\lambda \in \mathsf{P}$ which does not belong to the convex hull of $\{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}$. In fact, our vector λ belongs to a special 15-dimensional face F of P. Here is the definition of λ (below we abbreviate notation and write ab instead of $\{a,b\}$):

$$
\begin{aligned}
0 &= \lambda(\emptyset) = \lambda(a) = \lambda(b) = \lambda(c) = \lambda(d) = \lambda(ad) = \lambda(bc),\\
\frac{1}{2} &= \lambda(e) = \lambda(f)\\
&= \lambda(ae) = \lambda(af) = \lambda(be) = \lambda(bf) = \lambda(ce) = \lambda(cf) = \lambda(de) = \lambda(df)\\
&= \lambda(ab) = \lambda(ac) = \lambda(bd) = \lambda(cd)\\
&= \lambda(abe) = \lambda(acf) = \lambda(abd) = \lambda(acd) = \lambda(bcd),\\
1 &= \lambda(L) \qquad \text{for remaining } L \subseteq N.
\end{aligned}
$$

Let us denote by $\mathcal{Q}$ the class of those sets $L \subseteq N$ for which $\lambda(L) = 1$; note that $|\mathcal{Q}| = 38$. Observe that λ is non-decreasing: $\lambda(S) \leq \lambda(T)$ whenever $S \subseteq T \subseteq N$. It also belongs to the affine subspace A' of $\mathbb{R}^{\mathcal{P}(N)}$ specified by $45 = 7 + 38$ equalities from the first and last line above and 4 equalities

$$
\lambda^*(ab) = \lambda^*(abe),\ \lambda^*(ac) = \lambda^*(acf),\ \lambda^*(bd) = \lambda^*(abd),\ \lambda^*(cd) = \lambda^*(acd) \tag{11}
$$

required for $\lambda^* \in \mathbb{R}^{\mathcal{P}(N)}$. Let us put (we use our notational convention)

$$
\lambda' \;:=\; \frac{1}{2} \cdot \lambda_{\,|e|ab|bd|adf|bcf|cdf|} \;+\; \frac{1}{2} \cdot \lambda_{\,|f|ac|cd|ade|bce|bde|}
$$

and observe that

$$
\lambda = \lambda' - \frac{1}{2} \cdot \delta_{bcd} \qquad \text{where } \delta_{bcd} \text{ is the indicator of the set } bcd; \tag{12}
$$

this particular step is slightly tedious but straightforward. The verification of the inequalities (9) for λ breaks into 3 sub-cases.

- If G is an incomplete chordal graph over N which has a clique $K \in \mathcal{Q}$ then one can choose a junction tree for G in which K is a terminal clique and write the expression on the left-hand side of (9) in the form

$$
\underbrace{[\,\lambda(K) - \lambda(N)\,]}_{=1-1=0} \;+\; \sum_{C \in \mathcal{C}(G)\setminus\{K\}} \underbrace{[\,\lambda(C) - \lambda(S_C)\,]}_{\geq 0},
$$

where $S_C \in \mathcal{S}(G)$ with $S_C \subset C$ is assigned to C uniquely through the junction tree; see basic facts from Section 2.1. Thus, (9) follows from the fact that λ is non-decreasing.

The maximal sets in $\mathcal{P}(N)$ outside $\mathcal{Q}$ are sets from the set system

$$
\mathcal{K} = \{\, abd, abe, acd, acf, bcd, bf, ce, de, df \,\}.
$$

Thus, the other chordal graphs G over N have cliques that are subsets of sets in $\mathcal{K}$. In particular, the set bcd is not a separator in such a graph G because otherwise G would have a clique strictly containing bcd, which must be in $\mathcal{Q}$.

- If G is a chordal graph over N which has no clique in $\mathcal{Q} \cup \{bcd\}$ then, bcd is neither a clique nor a separator in G and, by (8) and (4), $\overline{\mathsf{m}}_G(bcd) = 0$. Hence, by (12), $\langle \overline{\mathsf{m}}_G, \lambda \rangle = \langle \overline{\mathsf{m}}_G, \lambda' \rangle$. Moreover, by Lemma 1, $\lambda' \in \mathsf{P}$. Hence, one has $\langle \overline{\mathsf{m}}_G, \lambda' \rangle \geq 0$, which implies the same inequality (9) for λ.

The remaining case is that G is a chordal graph over N which has no clique in $\mathcal{Q}$ while bcd is a clique in G. Since cliques of G are subsets of sets in $\mathcal{K}$, the options for cliques E of G containing e are $E \in \{e, ae, be, ce, de, abe\}$, while the options for cliques F of G containing f are $F \in \{f, af, bf, cf, df, acf\}$. Observe that abe and acf cannot be simultaneously cliques of G because then ab, ac, and bc are edges in G, which implies that abc is complete in G and contradicts the assumption that G has no clique in $\mathcal{Q}$. In particular, either one has $E \in \{e, ae, be, ce, de\}$ for any clique E of G containing e or one has $F \in \{f, af, bf, cf, df\}$ for any clique F of G containing f.

- If G is a chordal graph over N with $bcd \in \mathcal{C}(G)$ but without a clique in $\mathcal{Q}$ then one can choose a junction tree for G in which bcd is a terminal clique and write the expression on the left-hand side of (9) in the form

$$[\lambda(bcd) - \lambda(N)] + \sum_{C \in \mathcal{C}(G) \setminus \{bcd\}} [\lambda(C) - \lambda(S_C)],$$

  where $S_C \in \mathcal{S}(G)$ with $S_C \subset C$ assigned to C uniquely (by the choice of the junction tree). While $[\lambda(bcd) - \lambda(N)] = -\frac{1}{2}$ the terms $[\lambda(C) - \lambda(S_C)]$ are non-negative. Thus, to verify the validity of (9) it is enough to find at least one set $C \in \mathcal{C}(G) \setminus \{bcd\}$ for which $[\lambda(C) - \lambda(S_C)] = +\frac{1}{2}$. In the case that $E \in \{e, ae, be, ce, de\}$ for any clique E of G containing e one can take $C = E$ for one of these cliques: there is at least one assignment $E \mapsto S_E \in \mathcal{S}(G)$ with $e \notin S_E$, giving $\lambda(E) = \frac{1}{2}$ and $\lambda(S_E) = 0$. The same argument works when $F \in \{f, af, bf, cf, df\}$ for any clique F of G containing f. Hence, (9) holds for λ also in this case.

Altogether, one has $\lambda \in \mathsf{P}$. The fact that $\lambda \notin \mathrm{conv}(\{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P})$ can be verified by a contradiction. Otherwise, λ is a convex combination of some vectors $\lambda^* \in \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}$. Assume without loss of generality that λ is a convex combination of such vectors λ^* with all coefficients strictly positive. Note that λ belongs to $\mathsf{F} := \mathsf{P} \cap \mathsf{A}'$, which is a face of P (use Lemma 1 saying that every $\lambda' \in \mathsf{P}$ is non-decreasing). Hence, all these $\lambda^* \in \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}$ must belong to $\mathsf{F} \subseteq \mathsf{A}'$, that is, $\lambda^* \in \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{F}$. The next step is to show that the equalities defining A' imply that $\lambda^*(bcd) = 1$ for any $\lambda^* \in \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{F}$. Indeed, by Lemma 1, every such λ^* is the indicator of a singleton-containing filter $\mathcal{F}$ and it is enough to show that $bcd \in \mathcal{F}$:

- If $a \in \mathcal{F}$ then $abd \in \mathcal{F}$; by $\lambda^*(bd) = \lambda^*(abd)$ in (11), $bd \in \mathcal{F} \Rightarrow bcd \in \mathcal{F}$.
- If either $b \in \mathcal{F}$ or $c \in \mathcal{F}$ or $d \in \mathcal{F}$ then obviously $bcd \in \mathcal{F}$.
- If $e \in \mathcal{F}$ then $abe \in \mathcal{F}$ and, by $\lambda^*(ab) = \lambda^*(abe)$ in (11), $ab \in \mathcal{F} \Rightarrow abd \in \mathcal{F}$. By $\lambda^*(bd) = \lambda^*(abd)$ in (11) one observes $bd \in \mathcal{F} \Rightarrow bcd \in \mathcal{F}$.
- If $f \in \mathcal{F}$ then $acf \in \mathcal{F}$ and, by $\lambda^*(ac) = \lambda^*(acf)$ in (11), $ac \in \mathcal{F} \Rightarrow acd \in \mathcal{F}$. By $\lambda^*(cd) = \lambda^*(acd)$ in (11) one observes $cd \in \mathcal{F} \Rightarrow bcd \in \mathcal{F}$.

Since $\lambda^*(bcd) = 1$ holds for any $\lambda^* \in \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{F}$ the same must hold for their convex combination λ, which contradicts its definition saying $\lambda(bcd) = \frac{1}{2}$.

5. Conclusions

We have introduced the concept of a dual polyhedron to the chordal graph polytope D_N. The point is that (the characterization of) its vertices yield(s) a complete inequality description of D_N (Theorem 1). Our particular construction of such a dual polyhedron P allowed us to show that the original *chordal graph conjecture* is equivalent to the condition that all vertices of P are zero-one vectors (Theorem 2). This is indeed true in case $|N| \leq 5$; nevertheless, we have also showed that, in case $|N| = 6$, there is a vertex of P which is not a zero-one vector. Specifically, we have constructed a vector $\lambda \in \mathsf{P}$ which is not in the convex hull of zero-one vectors from P (see Example 7 in Section 4) and this fact already implies the existence of a non-zero-one vertex of P.

Thus, the original optimistic conjecture that facet-defining inequalities for the chordal graph polytope over N correspond to set systems, namely to certain clutters of subsets of N, has been refuted in case $|N| = 6$. Let us emphasize that this conclusion has been made on the basis of a theoretical heuristic analysis of the situation in case of $|N| = 6$, not by computing all the vertices of the dual polyhedron P in this case. Therefore, we don't know what are the vertices of P in this case and the question of what is the minimal set of inequalities delimiting the chordal graph polytope D_N in the case $|N| = 6$ remains open.

In particular, we don't know the number of facets for D_N when $|N| = 6$ and whether it exceeds the number of vertices or not. Recall that in cases $|N| \leq 5$, when facets are known to correspond to clutters of subsets of N, the number of facets of D_N is lower than or equal to the number of its vertices [12, § 3], and, thus, its inequality description is "simpler" than its vertex description. In such a situation the LP re-formulation of the respective optimization tasks is clearly beneficial in comparison with a direct brute-force optimization over vertices of D_N (= chordal graphs = statistical models).

This leads to a natural open geometric question whether, in the case $|N| \geq 6$, the number of facets of D_N is lower or not than the number of its vertices. This seems to be a question of substantial theoretical importance for the development of the ILP approach to learning of decomposable models based on characteristic imsets. If the answer is "yes" then it confirms the suitability of this particular ILP approach despite our refutal of the original chordal graph conjecture: then one should aim at a theoretical characterization of facet-defining inequalities of D_N in general. On the other hand, if the answer is "no" for $|N| = 6$ then this is a clear indication that research activity in this direction should be stopped. Hence, our proximate research effort will be directed to computationally enumerating vertices of the dual polyhedron P in the case $|N| = 6$ because this may answer the above geometric question when $|N| = 6$. Nevertheless, we are aware of the fact that this computational task may appear to infeasible (as it seems now).

Even if the idea of direct use of facets of the chordal graph polytope D_N in statistical learning decomposable models appears to be hopeless there is still another promising research direction. It is the idea of characterizing the edges of the chordal graph polytope D_N, which may appear to be a simpler theoretical task in comparison with the task of complete characterization of its facets. The (geometric) edges of a polytope are its faces of the dimension 1, each of them is a segment connecting two vertices of the polytope, but not any such segment is an edge. The point is that common LP optimization methods, in particular the *simplex method*, can be interpreted as search methods passing vertices of a polytope through its geometric edges. Thus, if edges of D_N are characterized theoretically then one can possibly design an optimization method inspired by the simplex method in which the maxima of common scores are searched in this way. Thus, another open question related to the concept of a chordal graph polytope which deserves attention is what are the geometric edges of this polytope. This follows a similar idea considered earlier in context of learning Bayesian network structure [28], where the structures were represented by the so-called essential graphs. An additional step towards geometric interpretation has been done recently in context of causal discovery [29], where edges were characterized for some faces of the respective *characteristic imset polytope* for learning Bayesian network structure.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research of Milan Studený and Václav Kratochvíl has been supported by the grant GAČR number 19-04579S. We would like to express our thanks to the reviewers of this paper for their useful comments.

Appendix A. Proof of Lemma 1

Let us recall what we are going to show.

Lemma 1. *Every $\lambda \in \mathsf{P}$ is non-decreasing: $\lambda(S) \leq \lambda(T)$ whenever $S \subseteq T$. In particular, the set $\mathsf{P} \subseteq [0,1]^{\mathcal{P}(N)}$ is a non-empty bounded polyhedron. One has $\lambda \in \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}$ iff λ is the indicator of a singleton-containing filter $\mathcal{F} \subseteq \mathcal{P}(N)$ with $\emptyset \notin \mathcal{F}$. Moreover, $\{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P} \subseteq \mathrm{ext}\,(\mathsf{P})$.*

Recall that $\mathrm{ext}\,(\mathsf{P})$ denotes the set of vertices (= extreme points) of P. In this proof, and also in other proofs presented in Appendices, we use the *italics* to distinguish those parts of the proof in which a subclaim/step is verified. We believe that this makes the proof better to follow for the reader.

Proof. Given $T \subset N$ and $i \in T$, consider a chordal graph G with cliques T and $N \setminus \{i\}$. The inequality (9) assigned to G then gives

$$\lambda(T) + \lambda(N \setminus \{i\}) - \lambda(T \setminus \{i\}) - \lambda(N) \geq 0\,.$$

Add the simple inequality $\lambda(N) - \lambda(N \setminus \{i\}) \geq 0$ to that to get $\lambda(T \setminus \{i\}) \leq \lambda(T)$. Using an inductive argument we observe that any $\lambda \in \mathsf{P}$ is non-decreasing.

The latter fact easily implies $\mathsf{P} \subseteq [0,1]^{\mathcal{P}(N)}$. As P is a bounded polyhedron, by [22, Corollary 8.7] it is a polytope and has finitely many vertices. Its non-emptiness follows from the next characterization of zero-one vectors in P.

Given $\lambda \in \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}$ we put $\mathcal{F} := \{T \subseteq N : \lambda(T) = 1\}$. Clearly, $\mathcal{F}$ is a filter. To see that $\mathcal{F}$ contains a singleton use (9) for the empty graph over N to observe $\sum_{i \in N} \lambda(\{i\}) \geq 1$. The constraint $\lambda(\emptyset) = 0$ implies $\emptyset \notin \mathcal{F}$.

Conversely, let λ be the indicator of a singleton-containing filter $\mathcal{F} \subseteq \mathcal{P}(N)$ with $\emptyset \notin \mathcal{F}$. The validity of simple monotonicity inequalities for λ is evident. To verify (9) for λ and some incomplete chordal graph G over N consider a junction tree $\mathcal{J}$ for G and introduce a sub-forest of $\mathcal{J}$ determined by $\mathcal{F}$: it has those hyper-nodes of $\mathcal{J}$ which belong to $\mathcal{F}$ and those hyper-edges in $\mathcal{J}$ which are labeled by sets from $\mathcal{F}$. The point is that the left hand side of (9) is equal to the number of connected components of the $\mathcal{F}$-sub-forest reduced by 1.

Indeed, recall that hyper-edges of $\mathcal{J}$ are labeled by separators and the expression $\sum_{C\in\mathcal{C}(G)}\lambda(C)-\sum_{S\in\mathcal{S}(G)}\nu_G(S)\cdot\lambda(S)$, which is the left-hand side of (9) plus 1, equals to the difference between the number of hyper-nodes of the $\mathcal{F}$-sub-forest and the number of its hyper-edges, which is just the number of its components.

The assumption that $\mathcal{F}$ contains a singleton implies that the $\mathcal{F}$-sub-forest has at least one hyper-node, and, thus, the left-hand side of (9) is non-negative.

The last claim follows from a simple geometric argument. Since P is a subset of the hypercube $[0,1]^{\mathcal{P}(N)}$, whose vertex set is $\{0,1\}^{\mathcal{P}(N)}$, every zero-one vector $\lambda\in\mathsf{P}$ is an extreme point of $[0,1]^{\mathcal{P}(N)}$ and, therefore, of P as well. □

Appendix B. Proof of Theorem 1

Let us recall what we are going to prove; note that $\mathrm{ext}(\mathsf{P})$ denotes the set of vertices of the dual polyhedron P. The re-writing of (10) below is based on the fact that $\kappa_\lambda(\emptyset)=0$ for any $\lambda\in\mathsf{P}$.

Theorem 1. *Assume $|N|\geq 2$. Given $\mathsf{c}\in\mathsf{A}$, one has $\mathsf{c}\in D_N$ if and only if $\mathsf{c}(N)\geq 0$ and the inequalities*

$$\sum_{\emptyset\neq S\subseteq N}\kappa_\lambda(S)\cdot\mathsf{c}(S)\geq 1 \tag{10}$$

hold for every $\lambda\in\mathrm{ext}(\mathsf{P})$.

In order to distinguish those parts of the proof(s) in which a second-order subclaim is verified (within the text in the *italics*) we indent those parts.

Proof. Let $\mathrm{cone}(\mathsf{P})$ denote the cone generated by P. Since P is a subset of the affine space specified by $\lambda(N)=1$ one has $\mathrm{cone}(\mathsf{P})=\mathrm{cone}(\mathrm{ext}(\mathsf{P}))$.

I. Observe that $\lambda\in\mathbb{R}^{\mathcal{P}(N)}$ belongs to $\mathrm{cone}(\mathsf{P})\equiv\mathrm{cone}(\mathrm{ext}(\mathsf{P}))$ if and only if $\lambda(\emptyset)=0$ and λ satisfies both the simple monotonicity inequalities and the graphical inequalities (9) for incomplete chordal graphs G over N.

Indeed, the necessity of those linear constraints easily follows from Definition 2. As concerns their sufficiency, by repeating the arguments in the 1st paragraph of the proof of Lemma 1 we observe that they imply that λ is non-decreasing; hence, $\lambda(\emptyset)=0$ implies $\lambda(N)\geq 0$ and λ is the $\lambda(N)$-multiple of a vector in P.

Let $\mathrm{lin}(\mathsf{S})$ denote the linear hull of a set $\mathsf{S}\subseteq\mathbb{R}^{\mathcal{P}(N)}$. Introduce for any $j\in N$ a vector $\lambda_{|j|}\in\mathbb{R}^{\mathcal{P}(N)}$ as the indicator of the filter generated by a trivial clutter $\mathcal{L}:=\{\{j\}\}$, that is, $\lambda_{|j|}(T)=\delta(j\in T)$ for any $T\subseteq N$.

II. Then we observe, for every $\lambda\in\mathbb{R}^{\mathcal{P}(N)}$ satisfying $\lambda(\emptyset)=0$, that one has

$$\lambda\in\mathrm{cone}\,(\,\mathrm{ext}(\mathsf{P})\cup\mathrm{lin}(\{\lambda_{|j|}:j\in N\})\,)$$

iff λ satisfies the inequalities (9) for all incomplete chordal graphs G over N.

Indeed, any $\lambda\in\mathrm{ext}(\mathsf{P})\subseteq\mathsf{P}$ satisfies inequalities (9) for all incomplete chordal graphs G over N. Since the conic combination preserves the validity of (9), their necessity follows from the fact that every $\lambda_{|j|}$ satisfies (9) with equality.

To this end one can repeat, for every $j\in N$, the arguments in the 4th paragraph of the proof of Lemma 1 with

$$\mathcal{F}_j:=\{T\subseteq N:\ j\in T\}$$

and realize that the $\mathcal{F}_j$-sub-forest of $\mathcal{J}$ has only one connected component (use the definition of a junction tree for this purpose).

For the sufficiency of the inequalities assume $\lambda\in\mathbb{R}^{\mathcal{P}(N)}$, $\lambda(\emptyset)=0$, satisfying (9) and put $\beta_j:=\lambda(N)-\lambda(N\setminus\{j\})$ for $j\in N$. Then $\lambda':=\lambda-\sum_{j\in N}\beta_j\cdot\lambda_{|j|}$ satisfied both $\lambda'(\emptyset)=0$ and the simple monotonicity inequalities with equality. Since λ' also satisfies all inequalities (9), by the previous Step I, $\lambda'\in\mathrm{cone}(\mathrm{ext}(\mathsf{P}))$, which allows us to observe that $\lambda\in\mathrm{cone}\,(\,\mathrm{ext}(\mathsf{P})\cup\mathrm{lin}(\{\lambda_{|j|}:j\in N\}))$.

One can interpret any $\lambda\in\mathbb{R}^{\mathcal{P}(N)}$ satisfying $\lambda(\emptyset)=0$ as a vector in $\mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$ and understand the inequalities (9) in this context (= ignore the component for the empty set because it plays no role). Analogous convention will concern the superset Möbius inversions m_G for incomplete graphs G over N and their modified versions $\overline{\mathsf{m}}_G$. The component for the empty set is then determined by other components and does not occur in any linear inequality of our interest.

III. The next step is to observe using a duality consideration that, for every $\overline{\mathsf{m}}\in\mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$, one has

$$\overline{\mathsf{m}}\in\mathrm{cone}\,(\,\{\overline{\mathsf{m}}_G:\ G\text{ is an incomplete chordal graph over }N\}\,)$$

iff $\sum_{T:j\in T}\overline{\mathsf{m}}(T)=0$ for any $j\in N$ and

$$\sum_{\emptyset \neq L \subseteq N} \overline{\mathsf{m}}(L) \cdot \lambda(L) \geq 0 \;\; \text{for any } \lambda \in \text{ext}\,(\mathsf{P}).$$

Consider two cones in $\mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$*: put*

$$\mathsf{K} := \text{cone}\,(\,\{\overline{\mathsf{m}}_G : G \text{ is an incomplete chordal graph over } N\}\,), \quad \textit{and}$$

$$\mathsf{L} := \text{cone}\,(\,\{+\lambda_{|j|} : j \in N\} \cup \{-\lambda_{|j|} : j \in N\} \cup \text{ext}\,(\mathsf{P})\,)\,.$$

By definition, they are both polyhedral cones, and, therefore, closed convex cones. Consider the scalar product in the space $\mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$ *defined by* $\langle \overline{\mathsf{m}}, \lambda \rangle := \sum_{\emptyset \neq S \subseteq N} \overline{\mathsf{m}}(S) \cdot \lambda(S)$*, which allows one to define the dual cone*

$$\mathsf{S}^* := \{\lambda \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}} \;:\; \langle \overline{\mathsf{m}}, \lambda \rangle \geq 0 \;\; \text{for all} \quad \overline{\mathsf{m}} \in \mathsf{S}\}$$

to every subset $\mathsf{S} \subseteq \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$*. Since* $\mathsf{L} = \text{cone}\,(\,\text{ext}\,(\mathsf{P}) \cup \text{lin}\,(\{\lambda_{|j|} : j \in N\})\,)$*, by Step* II*, one has* $\mathsf{L} = \mathsf{S}^*$ *for*

$$\mathsf{S} = \{\overline{\mathsf{m}}_G : G \text{ is an incomplete chordal graph over } N\}\,;$$

by the equality $\mathsf{S}^* = (\text{cone}\,(\mathsf{S}))^*$ *for any* $\mathsf{S} \subseteq \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$ *one then gets* $\mathsf{L} = \mathsf{K}^*$*. A well-known fact is that* $\mathsf{K} = \mathsf{K}^{**}$ *for any closed cone; see for example [30, Consequence 1]. In particular,* $\mathsf{L} = \mathsf{K}^*$ *gives* $\mathsf{L}^* = \mathsf{K}^{**} = \mathsf{K}$*. Again using* $(\text{cone}\,(\mathsf{S}))^* = \mathsf{S}^*$ *observe that* $\mathsf{K} = \mathsf{L}^*$ *consists of those* $\overline{\mathsf{m}} \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$ *for which* $\langle \overline{\mathsf{m}}, \lambda_{|j|} \rangle = 0$ *for any* $j \in N$ *and* $\langle \overline{\mathsf{m}}, \lambda \rangle \geq 0$ *for any* $\lambda \in \text{ext}\,(\mathsf{P})$*.*

This allows us to characterize the convex hull of our graphical vectors:

IV. for every $\overline{\mathsf{m}} \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$, one has

$$\overline{\mathsf{m}} \in \text{conv}\,(\,\{\overline{\mathsf{m}}_G : G \text{ is an incomplete chordal graph over } N\}\,)$$

iff $\overline{\mathsf{m}}(N) = -1$, $\sum_{T: j \in T} \overline{\mathsf{m}}(T) = 0$ for any $j \in N$ and

$$\sum_{\emptyset \neq L \subseteq N} \overline{\mathsf{m}}(L) \cdot \lambda(L) \geq 0 \quad \text{for any } \lambda \in \text{ext}\,(\mathsf{P}).$$

Indeed, by (8)*, one has* $\overline{\mathsf{m}}_G(N) = -1$ *for each incomplete chordal graph* G *over* N*, which implies, together with Step III, the necessity of the linear constraints. For their sufficiency use the other implication in Step III saying that* $\overline{\mathsf{m}}$ *belongs to the conic hull:* $\overline{\mathsf{m}} = \sum_k \alpha_k \cdot \overline{\mathsf{m}}_{G_k}$*, where* G_k *are (all) incomplete chordal graphs over* N *and* $\alpha_k \geq 0$*. The substitution* $-1 = \overline{\mathsf{m}}(N) = \sum_k \alpha_k \cdot \overline{\mathsf{m}}_{G_k}(N) = \sum_k \alpha_k \cdot (-1)$ *gives* $\sum_k \alpha_k = 1$*, that is,* $\overline{\mathsf{m}}$ *belongs to the convex hull.*

By definition, $\mathsf{m}_G(N) = 0$ for incomplete chordal graphs G; using (8) realize that $\mathsf{m}_G = \overline{\mathsf{m}}_G + \delta_N$, where $\delta_N \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$ denotes the indicator of N, and obtain from Step IV:

V. For every $\mathsf{m} \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$, one has

$$\mathsf{m} \in \text{conv}\,(\,\{\mathsf{m}_G : G \text{ is an incomplete chordal graph over } N\}\,)$$

iff $\mathsf{m}(N) = 0$, $\sum_{T: j \in T} \mathsf{m}(T) = 1$ for any $j \in N$ and

$$\sum_{\emptyset \neq L \subseteq N} \lambda(L) \cdot \mathsf{m}(L) \geq \lambda(N) = 1 \quad \text{for any } \lambda \in \text{ext}\,(\mathsf{P}).$$

Further re-writing is in terms of the vector c whose superset Möbius inversion is m. To this end use the inverse formula to (3) saying $\mathsf{c}(S) = \sum_{T: S \subseteq T} \mathsf{m}(T)$ for $S \subseteq N$ and obtain using the formula (6) the following:

VI. For every $\mathsf{c} \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$, one has

$$\mathsf{c} \in \text{conv}\,(\,\{\mathsf{c}_G : G \text{ is an incomplete chordal graph over } N\}\,)$$

iff $\mathsf{c}(N) = 0$, $\mathsf{c}(\{j\}) = 1$ for any $j \in N$ and

$$\sum_{\emptyset \neq S \subseteq N} \kappa_\lambda(S) \cdot \mathsf{c}(S) \geq 1 \quad \text{for any } \lambda \in \text{ext}\,(\mathsf{P}).$$

VII. Finally observe that, for any $\mathsf{c} \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$,

$$\mathsf{c} \in \text{conv}\,(\,\{\mathsf{c}_G : G \text{ is a chordal graph over } N\}\,)$$

iff $\mathsf{c}(N) \geq 0$, $\mathsf{c}(\{j\}) = 1$ for any $j \in N$ and

$$\sum_{\emptyset\neq S\subseteq N} \kappa_\lambda(S)\cdot \mathsf{c}(S)\geq 1 \quad \text{for any } \lambda\in \mathrm{ext}\,(\mathsf{P}).$$

Indeed, the characteristic imset c_H for the complete graph H over N clearly satisfies those linear constraints even with equalities except for $\mathsf{c}(N)\geq 0$: by (4) its superset Möbius inversion m_H is the indicator of the set N, use (6) and the fact that $\lambda(N)=1$ for every $\lambda\in \mathrm{ext}\,(\mathsf{P})\subseteq \mathsf{P}$. This, together with Step VI, implies the necessity of the linear constraints.
For the sufficiency of the considered linear constraints we first realize that they imply that c is non-increasing: $\mathsf{c}(S)\geq \mathsf{c}(T)$ whenever $\emptyset\neq S\subseteq T\subseteq N$.

To this end, for every fixed $\emptyset\neq S\subset N$ and $i\in N\setminus S$ consider the clutter $\mathcal{L}:=\{\{i\},S\}$ and the indicator $\lambda_{|i|S|}$ of the corresponding filter. By Lemma 1, $\lambda_{|i|S|}\in \mathrm{ext}\,(\mathsf{P})$. The inequality for $\lambda_{|i|S|}$, that is, the clutter inequality for $\mathcal{L}$, gives $\mathsf{c}(\{i\})+\mathsf{c}(S)-\mathsf{c}(\{i\}\cup S)\geq 1$ (see Section 2.3). Substitute $\mathsf{c}(\{i\})=1$ to get $\mathsf{c}(S)\geq \mathsf{c}(\{i\}\cup S)$; an inductive argument implies that c is non-increasing.

Because $\mathsf{c}(\{j\})=1$ for arbitrary $j\in N$, it implies $0\leq \mathsf{c}(N)\leq 1$. Let us put $\alpha:=\mathsf{c}(N)$. In case $\alpha=0$ we use directly Step VI to conclude that c is in the convex hull. In case of $\alpha=1$ use the fact that c is non-increasing to realize that $\mathsf{c}=\mathsf{c}_H$, which again implies the desired conclusion. In case $0<\alpha<1$ we put

$$\mathsf{c}':=\frac{1}{1-\alpha}\cdot(\mathsf{c}-\alpha\cdot\mathsf{c}_H),$$

which means that $\mathsf{c}=\alpha\cdot\mathsf{c}_H+(1-\alpha)\cdot\mathsf{c}'$ is a convex combination. It implies, for every $\lambda\in \mathrm{ext}\,(\mathsf{P})$, that

$$\begin{aligned}
&\sum_{\emptyset\neq S\subseteq N} \kappa_\lambda(S)\cdot \mathsf{c}'(S)\\
&=\frac{1}{1-\alpha}\cdot\Big(\underbrace{\sum_{\emptyset\neq S\subseteq N}\kappa_\lambda(S)\cdot\mathsf{c}(S)}_{\geq 1}-\alpha\cdot\underbrace{\sum_{\emptyset\neq S\subseteq N}\kappa_\lambda(S)\cdot\mathsf{c}_H(S)}_{=1}\Big)\\
&\geq\frac{1}{1-\alpha}\cdot(1-\alpha)=1\,.
\end{aligned}$$

Since one has both $\mathsf{c}'(N)=0$ and $\mathsf{c}'(\{j\})=1$ for any $j\in N$, by the previous Step VI, *c' is in the convex hull of c_G's for incomplete chordal graphs G over N. Hence, c is in the convex hull for all chordal graphs over N.*

One has $D_N=\mathrm{conv}\,(\{\mathsf{c}_G : G \text{ is a chordal graph over } N\})$ and $\mathsf{c}\in\mathsf{A}$ iff $\mathsf{c}(\{j\})=1$ for any $j\in N$. Thus, the last Step VII implies Theorem 1. □

Appendix C. Proof of Theorem 2

Let us recall what we are going to prove; note that we assume $|N|\geq 2$.

Theorem 2. *The chordal graph conjecture holds iff $\{0,1\}^{\mathcal{P}(N)}\cap\mathsf{P}=\mathrm{ext}\,(\mathsf{P})$.*

The steps of the proof for Theorem 2 correspond to the steps of the proof for Theorem 1; however, the order of the steps is inverse and $\mathrm{ext}\,(\mathsf{P})$ is replaced by $\{0,1\}^{\mathcal{P}(N)}\cap\mathsf{P}$ in the claims.

Proof. If $\{0,1\}^{\mathcal{P}(N)}\cap\mathsf{P}=\mathrm{ext}\,(\mathsf{P})$ then, by Theorem 1 and Lemma 1, one has $\mathsf{c}\in D_N$ iff $\mathsf{c}(N)\geq 0$ and $\sum_{S\subseteq N}\kappa_{\mathcal{L}}(S)\cdot\mathsf{c}(S)\geq 1$ for every singleton-containing clutter $\mathcal{L}\subseteq\mathcal{P}(N)$ (see Section 2.4). As mentioned earlier, the inequalities for clutters consisting of one (singleton) set only are superfluous. This implies the sufficiency for the condition from Conjecture 1.
We need to give a proof of the necessity of $\{0,1\}^{\mathcal{P}(N)}\cap\mathsf{P}=\mathrm{ext}\,(\mathsf{P})$ for the validity of Conjecture 1. Thus, assume that the conjecture holds, which means that $\mathsf{c}\in D_N$ iff $\mathsf{c}\in\mathsf{A}$, $\mathsf{c}(N)\geq 0$ and c satisfies the clutter inequalities for all singleton-containing clutters $\mathcal{L}\subseteq\mathcal{P}(N)$.

Indeed, recall from Section 2.3 that the inequality for a clutter consisting of one set holds with equality for $\mathsf{c}\in\mathsf{A}$.

As explained in Section 2.4, the clutter inequality for a singleton-containing clutter $\mathcal{L}\subseteq\mathcal{P}(N)$ takes the following form: $1\leq\sum_{\emptyset\neq S\subseteq N}\kappa_\lambda(S)\cdot\mathsf{c}(S)$, where λ is the indicator of the filter $\mathcal{L}^{\uparrow}$. By Lemma 1, λ is an indicator of singleton-containing filter $\mathcal{F}\subseteq\mathcal{P}(N)$ with $\emptyset\notin\mathcal{F}$ iff $\lambda\in\{0,1\}^{\mathcal{P}(N)}\cap\mathsf{P}$.

A. Hence, we observe that

$$\mathsf{c}\in \mathrm{conv}\,(\{\mathsf{c}_G : G \text{ is a chordal graph over } N\})$$

iff $\mathsf{c}(N) \geq 0$, $\mathsf{c}(\{j\}) = 1$ for any $j \in N$ and

$$\sum_{\emptyset \neq S \subseteq N} \kappa_\lambda(S) \cdot \mathsf{c}(S) \geq 1 \quad \text{for any } \lambda \in \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}.$$

B. The next step is to observe, for every $\mathsf{c} \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$, that

$$\mathsf{c} \in \text{conv}\,(\{\mathsf{c}_G : G \text{ is an incomplete chordal graph over } N\})$$

iff $\mathsf{c}(N) = 0$, $\mathsf{c}(\{j\}) = 1$ for any $j \in N$ and

$$\sum_{\emptyset \neq S \subseteq N} \kappa_\lambda(S) \cdot \mathsf{c}(S) \geq 1 \quad \text{for any } \lambda \in \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}.$$

Indeed, note that the only chordal graph H over N with $\mathsf{c}_H(N) > 0$ *is the complete graph, which implies the necessity of the considered linear constraints. Conversely, a well-known fact is that if* F' *is a face of polytope* $\mathsf{P}' = \text{conv}\,(\mathsf{S})$ *then* $\mathsf{F}' = \text{conv}\,(\mathsf{S} \cap \mathsf{F}')$*; in our case the face is defined by the equality* $\mathsf{c}(N) = 0$ *and the converse implication follows from Step A.*

One can use the inverse formula to (3) saying $\mathsf{c}(S) = \sum_{T:\,S\subseteq T} \mathsf{m}(T)$ for $S \subseteq N$ and (6) to re-write Step B in terms of the superset Möbius inversion:

C. For every $\mathsf{m} \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$, one has

$$\mathsf{m} \in \text{conv}\,(\{\mathsf{m}_G : G \text{ is an incomplete chordal graph over } N\})$$

iff $\mathsf{m}(N) = 0$, $\sum_{T:j\in T} \mathsf{m}(T) = 1$ for any $j \in N$ and

$$\sum_{\emptyset \neq L \subseteq N} \lambda(L) \cdot \mathsf{m}(L) \geq 1 \quad \text{for any } \lambda \in \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}.$$

Further re-writing is in terms of their modified versions given by (8); because $\mathsf{m}_G(N) = 0$ gives $\overline{\mathsf{m}}_G = \mathsf{m}_G - \delta_N$, where $\delta_N \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$ denotes the indicator of N, we obtain this:

D. For every $\overline{\mathsf{m}} \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$, one has

$$\overline{\mathsf{m}} \in \text{conv}\,(\{\overline{\mathsf{m}}_G : G \text{ is an incomplete chordal graph over } N\})$$

iff $\overline{\mathsf{m}}(N) = -1$, $\sum_{T:j\in T} \overline{\mathsf{m}}(T) = 0$ for any $j \in N$ and

$$\sum_{\emptyset \neq L \subseteq N} \lambda(L) \cdot \overline{\mathsf{m}}(L) \geq 0 \quad \text{for any } \lambda \in \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}.$$

The next step is to characterize the conic hull of that set of vectors:

E. For every $\overline{\mathsf{m}} \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$, one has

$$\overline{\mathsf{m}} \in \text{cone}\,(\{\overline{\mathsf{m}}_G : G \text{ is an incomplete chordal graph over } N\})$$

iff $\sum_{T:j\in T} \overline{\mathsf{m}}(T) = 0$ for any $j \in N$ and

$$\sum_{\emptyset \neq L \subseteq N} \lambda(L) \cdot \overline{\mathsf{m}}(L) \geq 0 \quad \text{for any } \lambda \in \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}.$$

The necessity of those linear constraints to $\overline{\mathsf{m}} \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$ *follows easily from Step* D. *For their sufficiency we need to evidence that they imply* $\overline{\mathsf{m}}(N) \leq 0$. *To this end we consider* $\overline{\mathsf{c}} \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$ *whose superset Möbius inversion is* $\overline{\mathsf{m}}$, *and show that* $\overline{\mathsf{c}}$ *is non-increasing.*

> *Indeed, owing to the inverse formula* $\overline{\mathsf{c}}(S) = \sum_{T:\,S\subseteq T} \overline{\mathsf{m}}(T)$ *for* $\emptyset \neq S \subseteq N$, *the constraints in terms of* $\overline{\mathsf{c}}$ *mean* $\overline{\mathsf{c}}(\{j\}) = 0$ *for* $j \in N$ *and* $\sum_{\emptyset \neq R \subseteq N} \kappa_\lambda(R) \cdot \overline{\mathsf{c}}(R) \geq 0$ *for any* $\lambda \in \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}$, *using* (6)*. Given* $\emptyset \neq S \subset N$ *and* $i \in N \setminus S$ *consider the clutter* $\mathcal{L} := \{\{i\}, S\}$ *and the indicator* $\lambda_{|i|S|} \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$ *of the corresponding filter. By Lemma 1,* $\lambda_{|i|S|} \in \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}$. *The substitution of* $\lambda_{|i|S|}$ *into the above inequality gives*
>
> $$\overline{\mathsf{c}}(\{i\}) + \overline{\mathsf{c}}(S) - \overline{\mathsf{c}}(\{i\} \cup S) \geq 0\,.$$
>
> *Because* $\overline{\mathsf{c}}(\{i\}) = 0$ *it gives* $\overline{\mathsf{c}}(S) \geq \overline{\mathsf{c}}(\{i\} \cup S)$ *and by an inductive argument observe that* $\overline{\mathsf{c}}$ *is non-increasing:* $\overline{\mathsf{c}}(S) \geq \overline{\mathsf{c}}(T)$ *whenever* $\emptyset \neq S \subseteq T \subseteq N$.

Since $\overline{c}(\{j\}) = 0$ for any $j \in N$ it gives $0 \geq \overline{c}(N) = \overline{m}(N)$.
In case $\overline{m}(N) = \overline{c}(N) = 0$ one necessarily has $\overline{c}(S) = 0$ for each $\emptyset \neq S \subseteq N$ and, thus, $\overline{m}(T) = 0$ for each $\emptyset \neq T \subseteq N$. Thus, $\overline{m}$ belongs to the conic hull then.
In case $\alpha := \overline{c}(N) = \overline{m}(N) < 0$ we introduce $\overline{m}'$ as $\frac{1}{|\alpha|}$-multiple of $\overline{m}$. Since $\overline{m}'(N) = -1$, by Step D, $\overline{m}'$ belongs to the convex hull of the respective set of vectors, which implies that $\overline{m}$ belongs to its conic hull.

F. The next step is to observe using a duality consideration that, for every $\lambda \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$, one has

$$\lambda \in \underbrace{\mathrm{cone}\,(\,\{+\lambda_{|j|} : j \in N\} \cup \{-\lambda_{|j|} : j \in N\} \cup (\{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P})\,)}_{\mathsf{R}}$$

iff λ satisfies the inequalities (9) for all incomplete chordal graphs G over N.
Indeed, let R be the conic hull of the vectors $\lambda \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$ written above. We, moreover, put

$$\mathsf{K} := \mathrm{cone}\,(\{\overline{m}_G : G \text{ is an incomplete chordal graph over } N\})\,.$$

The previous Step E implies that $\mathsf{K} = \mathsf{R}^*$. Hence, $\mathsf{K}^* = \mathsf{R}^{**} = \mathsf{R}$ because R is a closed cone; see [30, Consequence 1]. It remains to realize that K^* consists of those $\lambda \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$ that (9) is fulfilled, that is, $\langle \overline{m}_G, \lambda\rangle \geq 0$, holds for incomplete chordal graphs G over N: it follows from the equality $\mathsf{S}^* = (\mathrm{cone}\,(\mathsf{S}))^*$ valid for any $\mathsf{S} \subseteq \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$.

Recall our convention saying that any $\lambda \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$ is naturally extended to $\lambda \in \mathbb{R}^{\mathcal{P}(N)}$ by $\lambda(\emptyset) = 0$. A further observation is as follows:

G. For every $\lambda \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$, one has $\lambda \in \mathrm{cone}\,(\{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P})$ if and only if it satisfies the simple monotonicity inequalities $\lambda(N) - \lambda(N \setminus \{i\}) \geq 0$ for $i \in N$ and the inequalities (9) for all incomplete chordal graphs G over N.
Indeed, Definition 2 implies that any $\lambda \in \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}$ satisfies both the simple monotonicity inequalities and the inequalities (9). As the conic combination preserves their validity the necessity of the linear constraints is evident. For their sufficiency consider a vector $\lambda \in \mathbb{R}^{\mathcal{P}(N)\setminus\{\emptyset\}}$ satisfying all of them. By Step F

$$\lambda = \sum_{i \in N} \alpha_j \cdot \lambda_{|j|} + \lambda'$$

with $\alpha_j \in \mathbb{R}$ and λ' in the conic hull of $(\{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}) \setminus \{\lambda_{|j|} : j \in N\}$. Note that, by Lemma 1, elements of $\{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}$ are indicators of singleton-containing filters $\mathcal{F} \subseteq \mathcal{P}(N)$ with $\emptyset \notin \mathcal{F}$. Hence, the simple monotonicity inequality $\lambda(N) - \lambda(N \setminus \{j\}) \geq 0$ for $j \in N$ holds with equality for all elements of $\{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}$ except for $\lambda_{|j|}$. Thus, since the vector λ satisfies the simple monotonicity inequalities one has $\alpha_j \geq 0$ for $j \in N$. This implies that λ is in the conic hull of $\{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}$.

H. From that we observe that $\mathrm{conv}\,(\{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}) = \mathsf{P}$.
Indeed, $\mathrm{conv}\,(\{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}) \subseteq \mathsf{P}$ is evident. For converse inclusion consider $\lambda \in \mathsf{P}$ and, by Definition 2 and Step G, observe that $\lambda \in \mathrm{cone}\,(\{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P})$, that is, $\lambda = \sum_i \alpha_i \cdot \lambda_i$ with $\alpha_i \geq 0$ and $\lambda_i \in \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}$. Write $1 = \lambda(N) = \sum_i \alpha_i \cdot \lambda_i(N) = \sum_i \alpha_i$, which implies that it is a convex combination.

The last step implies that $\mathrm{ext}\,(\mathsf{P}) \subseteq \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}$, which together with the converse inclusion from Lemma 1 gives the observation $\mathrm{ext}\,(\mathsf{P}) = \{0,1\}^{\mathcal{P}(N)} \cap \mathsf{P}$, which is the desired conclusion. Thus, the necessity of this condition for the validity of Conjecture 1 has been verified. □

References

[1] S.L. Lauritzen, Graphical Models, Clarendon Press, Oxford, 1996.
[2] R.G. Cowell, A.P. Dawid, S.L. Lauritzen, D.J. Spiegelhalter, Probabilistic Networks and Expert Systems, Springer, New York, 1999.
[3] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufman, San Mateo, 1988.
[4] R.E. Neapolitan, Learning Bayesian Networks, Pearson Prentice Hall, Upper Saddle River, 2004.
[5] G.E. Schwarz, Estimation of the dimension of a model, Ann. Stat. 6 (1978) 461–464.
[6] D. Heckerman, D. Geiger, D.M. Chickering, Learning Bayesian networks: the combination of knowledge and statistical data, Mach. Learn. 20 (1995) 194–243.
[7] J. Cussens, M. Studený, D. Haws, Polyhedral aspects of score equivalence in Bayesian network structure learning, Math. Program., Ser. A 164 (1/2) (2017) 285–324.
[8] R. Hemmecke, S. Lindner, M. Studený, Characteristic imsets for learning Bayesian network structure, Int. J. Approx. Reason. 53 (2012) 1336–1349.
[9] M. Studený, D. Haws, Learning Bayesian network structure: towards the essential graph by integer linear programming tools, Int. J. Approx. Reason. 55 (2014) 1043–1071.
[10] M. Studený, J. Cussens, The chordal graph polytope for learning decomposable models, in: A. Antonucci, G. Corani, C.P. de Campos (Eds.), Proceedings of Machine Learning Research, vol. 52, PGM 2016, 2016, pp. 499–510.
[11] S. Lindner, Discrete optimization in machine learning: learning Bayesian network structures and conditional independence implication, Ph.D. thesis, TU Munich, 2012.
[12] M. Studený, J. Cussens, Towards using the chordal graph polytope in learning decomposable models, Int. J. Approx. Reason. 88 (2017) 259–281.

[13] M. Studený, J. Cussens, V. Kratochvíl, Dual formulation of the chordal graph conjecture, in: M. Jaeger, T.D. Nielsen (Eds.), Proceedings of Machine Learning Research, vol. 138, PGM 2020, 2020, pp. 449–460.
[14] K.S. Sesh Kumar, F. Bach, Convex relaxations for learning bounded-treewidth decomposable graphs, in: S. Dasgupta, D. McAlester (Eds.), Proceedings of Machine Learning Research 28, vol. 1, ICML 2013, 2013, pp. 525–533.
[15] A. Pérez, C. Blum, J.A. Lozano, Learning maximum weighted $(k+1)$-order decomposable graphs by integer linear programming, in: L.C. van der Gaag, A.J. Feelders (Eds.), PGM 2014, in: Lecture Notes in AI, vol. 8754, 2014, pp. 396–408.
[16] A. Pérez, C. Blum, J.A. Lozano, Approximating the maximum weighted decomposable graph problem with applications to probabilistic graphical models, in: V. Kratochvíl, M. Studený (Eds.), Proceedings of Machine Learning Research, vol. 72, PGM 2018, 2018, pp. 320–331.
[17] J. Corander, T. Janhunen, J. Rintanen, H. Nyman, J. Pensar, Learning chordal Markov networks by constraint satisfaction, in: C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K.Q. Weiberger (Eds.), Advances in Neural Information Processing Systems, vol. 26, Curran Associates, 2013, pp. 1349–1357.
[18] K. Kangas, T. Niinimäki, M. Koivisto, Learning chordal Markov networks by dynamic programming, in: Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems, vol. 27, Curran Associates, 2014, pp. 2357–2365.
[19] K. Kangas, T. Niinimäki, M. Koivisto, Averaging decomposable graphs by dynamic programming and sampling, in: M. Meila, T. Heskes (Eds.), Uncertainty in Artificial Intelligence, vol. 31, AUAI Press, 2015, pp. 415–424.
[20] K. Rantanen, A. Hyttinen, M. Järvisalo, Learning chordal Markov networks via branch and bound, in: I. Guyon, U. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, vol. 30, 2017, pp. 1845–1855.
[21] M. Studený, Probabilistic Conditional Independence Structures, Springer, London, 2005.
[22] A. Barvinok, A Course in Convexity, American Mathematical Society, Providence, 2002.
[23] V. Chvátal, Linear Programming, W.H. Freeman, New York, 1983.
[24] L.A. Wolsey, Integer Programming, John Wiley, New York, 1998.
[25] R.P. Stanley, Enumerative Combinatorics, vol. 1, Cambridge University Press, Cambridge, 1997.
[26] A. Brøndsted, An Introduction to Convex Polytopes, Springer, New York, 1983.
[27] G.M. Ziegler, Lectures on Polytopes, Springer, New York, 1995.
[28] M. Studený, J. Vomlel, R. Hemmecke, A geometric view on learning Bayesian network structures, Int. J. Approx. Reason. 51 (2010) 573–586.
[29] S. Linusson, P. Restadh, L. Solus, Greedy causal discovery is geometric, arxiv paper, https://arxiv.org/abs/2103.03771, March 2021.
[30] M. Studený, Convex cones in finite-dimensional real vector spaces, Kybernetika 29 (2) (1993) 180–200.