Embedding Formulations and Complexity for Unions of Polyhedra

Juan Pablo Vielma
Sloan School of Management, Massachusetts Institute of Technology, jvielma@mit.edu

Mixed integer programming formulations for unions of polyhedra or (polyhedral) disjunctive constraints can be divided into extended formulations that use both 0-1 and continuous auxiliary variables, and non-extended formulations that only use the 0-1 variables that are strictly necessary to construct a valid formulation. Standard extended formulations have sizes that are linear on the sizes of the polyhedra and have linear programming relaxations with extreme points that naturally satisfy the appropriate integrality constraints (such formulations are usually denoted ideal and are as strong as possible). In contrast, for non-extended formulation there usually is an important trade-off between strength and size. However, a flexible use of the 0-1 variables that signal the selection among the polyhedra can lead to non-extended formulations that are ideal and smaller than the best extended alternative. Furthermore, these formulations have been shown to provide a significant computation advantage. This paper attempts to explain and expand the success of such formulations by introducing a geometric characterization of them. This characterization is based on an embedding of the polyhedra in a higher dimensional space and provides a systematic procedure to construct ideal formulations. Furthermore, the characterization naturally leads to two notions of formulation complexity for unions of polyhedra: 1) the size of the smallest non-extended formulation, and 2) the size of the smallest ideal non-extended formulations. We analyze such measures for various disjunctions of practical interest by providing (nearly) matching upper and lower bounds, with special emphasis on the number of general inequalities (i.e. those that are not variable bounds). In particular, when analyzing Special Order Sets of type 2 (SOS2) we show optimality with respect to the number of general inequalities of an existing ideal formulation and introduce a non-ideal formulation that only uses a constant number of of general inequalities. Finally, we consider how adding redundancy to the disjunction can simplify the construction of ideal formulations and compare the formulation complexity measures to other complexity notions such as the size of the convex hull of the union of the polyhedra and the extension complexity of this convex hull.

Key words:
MSC2000 subject classification:
OR/MS subject classification: Primary: ; secondary:
History: June 2nd, 2015

1. Introduction
Several techniques in combinatorial optimization can be interpreted as the construction of a strong linear programming (LP) or integer programming (IP) formulation of a discrete set. Indeed, most combinatorial optimization problems can be described as

\[
\begin{align*}
\text{max} & \quad \sum_{i=1}^{d} c_{i}x_{i} \\
\text{s.t.} & \quad x \in S,
\end{align*}
\]

(1a)

(1b)

where \( c \in \mathbb{R}^{d} \) and \( S \subseteq \{0,1\}^{d} \) is an implicitly defined discrete set (e.g. \( S \) could be the set of incidence vectors of matchings in a graph). While such problems can be solved directly with specialized algorithms of various levels of theoretical and practical efficiency, an extremely versatile
and effective solution method arises from the observation that (1) is equivalent to

\[
\begin{align*}
\text{max} & \quad \sum_{i=1}^{d} c_i x_i \\
\text{s.t.} & \quad x \in Q, \\
& \quad x \in \{0, 1\}^d,
\end{align*}
\]

(2a)

(2b)

(2c)

where \(Q \subseteq \mathbb{R}^d\) is any polyhedron such that

\[
Q \cap \{0, 1\}^d = S.
\]

(3)

Together with integrality requirements (2c), any polyhedron \(Q\) that satisfies (3) yields a pure 0-1 IP formulation (i.e. one that does not include general integer or continuous variables) of \(S\) and in turn the corresponding (2) yields a pure 0-1 IP formulation of the associated combinatorial optimization problem. If \(Q\) is equal to the convex hull \(\text{conv}(S)\) of \(S\), then the IP formulation is often denoted perfect or integral. In such case, (2c) can be removed and formulation (2) is actually an LP. If in addition \(Q\) is polynomial sized or has an efficient description\(^\dagger\), then both (1) and (2) can be solved in polynomial time with an appropriate LP algorithm. Determining the existence of such efficiently described integral formulations and giving explicit constructions of them are the basis of extremely prolific areas of theoretical development including polyhedral combinatorics [39, 43, 44] and extended formulations for combinatorial optimization [12, 29, 57]. Similarly, while finding such efficient descriptions of \(\text{conv}(S)\) for NP-hard problems is impossible unless \(P=NP\), constructing and understanding the structure of polyhedra that are relatively close to \(\text{conv}(S)\) (i.e. constructing nearly-integral formulations) is the basis of many approximation algorithms and practical methods for such intractable combinatorial optimization problems [4, 37, 49, 56]. In addition, the techniques used to construct such integral/nearly-integral and efficiently described formulations can result in extremely effective solution methods for problems in which \(x \in S\) is only one of many constraints.

For instance, strong formulations for the Traveling Salesman Problem can be useful for Vehicle Routing problems [48] and strong formulations for the Spanning and Steiner Tree Problems can be useful for problems that include connectivity constraints [10]. Because of this, there has recently been a surge of publications studying techniques for constructing or showing the impossibility of constructing polynomial sized integral formulations for discrete sets associated to combinatorial optimization problems (e.g. [2, 21, 22, 23, 30, 31, 41, 42]). However, while discrete sets are extremely versatile, their modeling power is still limited when compared to mixed integer programming (MIP) formulations that allow both discrete and continuous variables. For this reason, in this paper we study techniques for finding the smallest size of a MIP formulation of disjunctive constraints of the form

\[
x \in \bigcup_{i=1}^{n} P^i
\]

(4)

where \(\{P^i\}_{i=1}^{n}\) is a finite family of polyhedra in \(\mathbb{R}^d\), which share a common recession cone.

In 1984 Jeroslow and Lowe showed that disjunctive constraints of the form (4) are precisely what can be modeled using linear MIP formulations with bounded or 0-1 integer variables [28, 35]. Since then, several formulation techniques for such constraints have been developed and applied to a wide range of problems (e.g. see [50] and the references within). Some recent applications that

\(^\dagger\)i.e. if \(Q\) has a polynomial number of inequalities with polynomial sized coefficients, \(Q\) is the projection of a polyhedron with a polynomial number of inequalities with polynomial sized coefficients or the inequalities of \(Q\) can be separated in polynomial time.
Polynomially sized MIP formulations for (4) can be divided into two classes depending on their strength and types of auxiliary variables. The first class corresponds to strong extended formulations that use both 0-1 and continuous auxiliary variables (e.g. [50, Section 5]). This class includes the classical formulations by Balas, Jeroslow and Lowe, which have LP relaxations whose extreme points naturally satisfy the appropriate integrality constraints. Formulations, with such characteristic are as strong as possible and are usually denoted ideal and are the mixed-integer analog of integral formulations for discrete sets. The second class corresponds to non-extended formulations that exclude continuous auxiliary variables and only use the 0-1 variables that are strictly needed to construct a valid formulation for (4) (e.g. [50, Section 6]). This class includes traditional Big-M formulations and ad-hoc formulations for special structures that often fail to be ideal, but may still be relatively strong under certain conditions (e.g. [5, 8, 27]). A common feature of both classes is the use of \( n \) 0-1 variables that are constrained to add up to one to signal the selection of among the \( n \) polyhedra in the disjunction. In such setting, the use of continuous auxiliary variables can be a necessary condition for constructing ideal polynomial formulations (e.g. [50, Section 4]). However, different uses of 0-1 variables can lead to polynomial sized non-extended formulations that are ideal and smaller than standard extended formulations [52]. Furthermore, the resulting formulations can provide a significant computational advantage [52, 51]. This paper attempts to explain and expand the success of such formulations through the following contributions:

1. **Geometric characterization of flexible non-extended formulations:** Geometric characterization (3) of a pure IP formulation for a discrete set is extremely useful to analyze formulation strength and evaluate the size of the smallest formulation. Unfortunately, a similar characterization cannot be constructed for disjunctive constraint (4) in its \( d \)-dimensional ambient space (i.e. the space of the \( x \) variables). For this reason we consider an embedding of the constraint into a higher dimensional space that includes both the space of the \( x \) variables and the space of the 0-1 variables needed to construct a valid formulation. This embedding allows for an extension of the geometric characterization to the mixed-integer setting and, in particular, yields a systematic geometric construction of ideal non-extended formulations. Furthermore, the flexible use of 0-1 variables can be automatically considered through alternate embeddings of the constraint into different higher dimensional spaces.

2. **Bounds on formulation sizes and complexity of unions of polyhedra:** The geometric characterization of non-extended formulations naturally leads to two notions of complexity for unions of polyhedra: 1) the size of the smallest non-extended formulation, and 2) the size of the smallest ideal non-extended formulations. Constructing explicit ideal or non-ideal formulation yield upper bounds on these complexity measures and is clearly of practical interest. However, as highlighted by the recent activity on extended formulations for discrete sets, it is also important to give lower bounds on these complexity measures to evaluate how far current formulations are from the smallest possible formulation. Showing these lower bounds is complicated by the need to consider every possible embedding of the associated disjunctive constraints. However, through various techniques we are able to give (nearly) matching upper and lower bounds for various classes of disjunctions of practical interest. In particular, we show optimality with respect to the number of general inequalities (i.e. those that are not variable bounds) for some ideal formulations introduced in [52]. We also show the somewhat surprising result that the extensively studied Special Order Sets of type 2 (SOS2) constraints introduced by Beale and Tomlin in 1970 [7] can be formulated with a constant (independent of the natural size of the constraint) number of general inequalities. To the best of our knowledge, these results are the first lower bounds on sizes of MIP formulations.
3. Practical formulation constructions through redundancy: Constructing the smallest ideal formulation of a disjunction requires significant understanding of the geometry of the disjunction. For this reason it is also interesting to develop formulation techniques that can yield small formulations with less detailed knowledge of the structure of the constraints. One such techniques was introduced in [52] to construct formulations for some specially structured piecewise linear functions of two variables. We show how this technique fits in the geometric characterization of non-extended formulations and how it can be extended to a wider class of piecewise linear functions. In particular, we show that adding some redundancy to the disjunctive formulation can significantly reduce the formulation complexity and may be necessary to obtain the smallest possible formulation.

Throughout the paper we use the following notation. For a set $S \subseteq \mathbb{R}^d$ we let $\text{conv} (S)$, $\text{aff} (S)$, $\text{span} (S)$ and $\text{dim} (S)$ be the convex hull, affine hull, linear span and the dimension of $S$ respectively. For a polyhedron $P \subseteq \mathbb{R}^d$ we let $\text{ext} (P)$ and $\text{ray} (P)$ be the set of extreme points and extreme rays of $P$. We also let $P_\infty$ be the recession cone of $P$. Given two vectors $a, b \in \mathbb{R}^V$ for a finite index set $V$ we let $a \cdot b = \sum_{v \in V} a_v b_v$ be the inner product between $a$ and $b$. We also let $0 \in \mathbb{R}^V$ be the vector of all zeros and $e_u^v \in \mathbb{R}^V$ be the unit vector such that $e_u^v = 1$ if $u = v$ and $e_u^v = 0$ otherwise. Finally, we let $[n] := \{1, \ldots, n\}$ and $[a, b] := \{a, a+1, \ldots, b-1, b\}$.

The rest of the paper is organized as follows. In Section 2 we introduce the formulation characterizations and construction procedure based on embedding the polyhedra in a higher dimensional space. We also introduce the associated complexity measures and present an initial discussion about their relation to other natural measures. We also motivate the need for the flexible use of 0-1 variables. In Section 3 we introduce a special class of disjunctions that will be the focus of most of the paper. This class greatly simplifies our analysis and already has a wide range of applications. We end Section 3 by considering basic properties of the formulations for this class of disjunctions. Then, in Section 4 we study the sizes of the smallest non-extended formulation for the same class of disjunctions and in Section 5 we study the size of the smallest ideal non-extended formulation for a specific member of this class. Section 6 considers a variant of the proposed formulation approach that can be more amenable to practical constructions. Finally, Section 7 considers further relations between the complexity measures and discusses some possible directions of future research.

2. Formulations and Complexity Consider a discrete set $S \subseteq \mathbb{Z}^d$ such that $\text{conv} (S)$ is a polyhedron and $S = \text{conv} (S) \cap \mathbb{Z}^d$. A geometric procedure to construct an IP formulation for $x \in S$ that does not utilize auxiliary variables is through a polyhedron $Q \subseteq \mathbb{R}^d$ such that $S = Q \cap \mathbb{Z}^d$ (e.g. [31]). For such $Q$, an IP formulation of $x \in S$ is simply given by $x \in Q \cap \mathbb{Z}^d$ and $Q$ is the LP relaxation of the formulation. In this context the strongest possible formulation is $Q = \text{conv} (S)$, which has integral extreme points. Such strongest formulation is usually denoted an integral formulation. In contrast, consider a disjunctive constraint of the form

$$x \in \bigcup_{i=1}^{n} P^i$$

where $\mathcal{P} := \{P^i\}_{i=1}^{n}$ is a finite family of non-empty polyhedra in $\mathbb{R}^d$, which for simplicity we initially assume are bounded. Constructing a MIP formulation for (5) cannot be done through the geometric procedure described above as there might not be any polyhedron $Q \subseteq \mathbb{R}^d$ such that $\bigcup_{i=1}^{n} P^i = Q \cap \mathbb{Z}^d$. Indeed, constructing a MIP formulation for (5) requires the use of at least some integer constrained auxiliary variables. For instance, most standard MIP formulations for (5) use 0-1 variables $y \in \{0,1\}^n$ such that $\sum_{i=1}^{n} y_i = 1$ (e.g. [50, Section 6]). These 0-1 variables indicate the selection among the $n$ polyhedra so that $x \in P^i$ whenever $(x,y) \in Q$ and $y = e^i$. A geometric interpretation of such formulations that is analogous to the discrete setting is
obtained if we embed the polyhedra into a space that also contains the 0-1 variables to obtain the disjunctive constraint given by
\[(x, y) \in \bigcup_{i=1}^{n} (P_i \times \{e^i\}).\] (6)

Then a formulation of (6) is given by a polyhedron \(Q \subseteq \mathbb{R}^{d+n}\) such that \(\bigcup_{i=1}^{n} (P_i \times \{e^i\}) = Q \cap (\mathbb{R}^d \times \mathbb{Z}^n)\) and the strongest possible formulation is given by \(Q = \text{conv} \left( \bigcup_{i=1}^{n} (P_i \times \{e^i\}) \right)\). This embedding of \(\bigcup_{i=1}^{n} P_i\) in a higher dimensional space is known as the Cayley trick where either \(\bigcup_{i=1}^{n} (P_i \times \{e^i\})\) or \(\text{conv} \left( \bigcup_{i=1}^{n} (P_i \times \{e^i\}) \right)\) is sometimes denoted the Cayley Embedding (e.g. [26, 32, 54]). A natural generalization of the Cayley Embedding will be the basis for constructing small and strong formulations that do not use auxiliary variables besides the 0-1 variable that are strictly necessary to model disjunction (5).

The traditional use of 0-1 variables that add to one can be interpreted as a unary encoding of the selection among the polyhedra. However, one key to construct small formulations for unions of polyhedra is a flexible use of 0-1 variables that go beyond this unary encoding. This encoding flexibility can be formalized through the following definition.

**Definition 1 (Encoding).** For \(n, k \in \mathbb{Z}_+\) let
\[H_k(n) := \left\{ \{h^i\}_{i=1}^{n} \subseteq \{0,1\}^k : h^i \neq h^j \quad \forall i \neq j \right\}\]
be the family of all encodings of \(n\) alternatives with \(k\) bits or 0-1 variables and
\[H(n) := \bigcup_{k \in \mathbb{Z}_+} H_k(n)\]
be the family of all possible encodings of \(n\) alternatives.

**Example 1.** The encoding associated with traditional formulations corresponds to \(k = n\) and \(h^i = e^i\) for all \(i \in [n]\), which we refer to as the unary encoding.

An alternative family of encodings that appears in logarithmic formulations (e.g. [50, Section 9]), utilizes \(k = \lfloor \log_2 n \rfloor\). This family of encodings are the ones with the smallest possible number of variables and, when \(n\) is a power of 2, they satisfy \(\{h^i\}_{i=1}^{n} = \{0,1\}^k\). We refer to this family as the binary encodings. The only difference between the members of the binary encoding family for \(n\) a power of 2, is the specific ordering of the vectors in the encoding. Together with the specific ordering of the polyhedra in \(\mathcal{P}\), this induces a specific vector-polyhedron assignment, which can have a significant impact on formulation sizes (e.g. see Section 5). Similar alternative assignments for the unary encoding can be obtained by changing the order in \(\mathcal{P}\). However, in Lemma 1 we will see that these alternate assignments are inconsequential, which justifies our consideration of a unique unary encoding.

Using these abstract encodings we get the following definition of a formulation, which uses the generalization of embedding (6) given by \((x, y) \in \bigcup_{i=1}^{n} (P_i \times \{e^i\})\).

**Definition 2 (Formulation).** Let \(\mathcal{P} := \{P_i\}_{i=1}^{n}\) be a finite family of non-empty polyhedra in \(\mathbb{R}^d\), \(H := \{h^i\}_{i=1}^{n} \in H_k(n)\) and \(Q \subseteq \mathbb{R}^{d+k}\) be a polyhedron. We say \((Q, H)\) is a formulation for \(\bigcup_{i=1}^{n} P_i\) if and only if
\[(x, h) \in Q \cap (\mathbb{R}^d \times \mathbb{Z}^k) \iff \exists i \in [n] \text{ s.t. } h = h^i \land x \in P^i\]

The only auxiliary variables used by the formulation in Definition 2 are the 0-1 variables that are strictly necessary to construct a valid formulation. Such formulations are sometimes denoted projected formulations (e.g. [50, Section 6]). However, for simplicity we here denote them just formulations and refer to formulations that may use additional auxiliary variables as extended formulations (see Definition 8). Similarly, while Definition 2 already uses a generalization of the
Cayley Embedding we reserve the term embedding formulation for the strongest possible formulation for a given encoding (see Definition 5). Before introducing such strongest formulation we review two standard measures of formulation strength (e.g. [50, Section 2.2]).

The first measure of formulation strength considers how close the LP relaxation of the formulation (in our case Q) is to the convex hull of the set we are trying to formulate. The following definition formalizes this for unions of polyhedra and expands the concept to strength with regards to the encoding.

**Definition 3 (Sharp and Encoding-Sharp Formulations).** Let (Q, H) be a formulation for $\bigcup_{i=1}^{n} P^i$. We say (Q, H) is sharp if and only if

$$
\text{proj}_x (Q) := \{ x \in \mathbb{R}^d : \exists y \in \mathbb{R}^k \text{ s.t. } (x, y) \in Q \} = \text{conv} \left( \bigcup_{i=1}^{n} P_i \right)
$$

and that it is encoding-sharp if and only if

$$
\text{proj}_y (Q) := \{ y \in \mathbb{R}^k : \exists x \in \mathbb{R}^d \text{ s.t. } (x, y) \in Q \} = \text{conv} (H).
$$

An even stronger measure of strength is to require all extreme points of the LP relaxation to satisfy the encoding.

**Definition 4 (Ideal Formulation).** Let (Q, H) be a formulation for $\bigcup_{i=1}^{n} P^i$. We say (Q, H) is ideal if and only if

$$
\text{ext} (Q) \subseteq \mathbb{R}^d \times \{ 0, 1 \}^k.
$$

Ideal formulations are the natural mixed integer analog of an integral formulation for a set of integer points. Ideal formulations are certainly sharp and encoding sharp [50], but there are sharp and encoding sharp formulations that are not ideal (e.g. Examples 4 and 5). One systematic way to construct ideal formulations for unions of polyhedra is given in the following definition.

**Definition 5 (Embedding Formulation).** Let $\mathcal{P} := \{ P^i \}_{i=1}^{n}$ be a finite family of non-empty polyhedra in $\mathbb{R}^d$ and $H := \{ h^i \}_{i=1}^{n} \in \mathcal{H}_k(n)$. We define

$$
Q (\mathcal{P}, H) := \text{conv} \left( \bigcup_{i=1}^{n} P^i \times \{ h^i \} \right).
$$

Proposition 1 below shows that $Q (\mathcal{P}, H)$ is essentially the unique ideal formulation for a given encoding and hence we could refer to $Q (\mathcal{P}, H)$ as the ideal embedding formulation for encoding H. However, for simplicity we just refer to it as the embedding formulation for encoding H. If clear from the context we use $Q (\mathcal{P}, H)$ to refer both to the formulation and its LP relaxation.

**Proposition 1.** If $(Q^0, H)$ is an ideal formulation of $\bigcup_{i=1}^{n} P^i$ and $Q^0$ is a rational polyhedron, then $Q^0 = Q (\mathcal{P}, H)$.

Conversely, if $P^i_\infty = P^j_\infty$ for all $i, j \in [n]$ then $Q (\mathcal{P}, H)$ is a polyhedron and $(Q (\mathcal{P}, H), H)$ is an ideal formulation for $\bigcup_{i=1}^{n} P^i$.

**Proof.** For the first statement note that under the assumptions on $Q^0$ we have $Q^0_\infty = P^i_\infty \times \{ 0 \}$ for all $i \in [n]$ and hence all $P^i$ have the same recession cone. Hence, without loss of generality,

$$
Q^0 = \text{conv} \left( \bigcup_{i=1}^{n} \bigcup_{v \in V^i} \{ v \times h^i \} \right) + \text{cone} \left( \bigcup_{r \in \text{ray}(P^1)} \{ r \times 0 \} \right)
$$

where ray$(P^1)$ is the set of extreme rays of $P^1$ and $V^i \subseteq P^i$ and $|V^i| < \infty$ for all $i \in [n]$. 

Now, under the assumption $P^i_{\infty} = P^j_{\infty}$ for all $i,j \in [n]$, $\{P^i \times \{h^i\}\}_{i=1}^n$ is a finite family of non-empty polyhedra with identical recession cones. Then, by Lemma 4.41 and Corollary 4.44 in [13], $Q(\mathcal{P}, H)$ is a polyhedron and

$$Q(\mathcal{P}, H) = \text{conv} \left( \bigcup_{i=1}^n \bigcup_{v \in \text{ext}(P^i)} \{v \times h^i\} \right) + \text{cone} \left( \bigcup_{r \in \text{ray}(P^1)} \{r \times \mathbf{0}\} \right). \quad (7)$$

Hence, $Q^0 \subseteq Q(\mathcal{P}, H)$. The reverse inclusion follows by noting that, if $(Q^0, H)$ is a formulation, then $\bigcup_{i=1}^n P^i \times \{h^i\} \subseteq Q^0$.

For the second statement we again have that $Q(\mathcal{P}, H)$ is a polyhedron and (7) holds. Then, because the $h^i$'s are distinct extreme points of $[0,1]^k$, we have that if $(x,h) \in Q(\mathcal{P}, H) \cap (\mathbb{R}^d \times \mathbb{Z}^k)$ then $h = h^i$ for some $i \in [n]$ and

$$(x,h) \in \text{conv} \left( \bigcup_{v \in \text{ext}(P^i)} \{v \times h^i\} \right) + \text{cone} \left( \bigcup_{r \in \text{ray}(P^1)} \{r \times \mathbf{0}\} \right) = P^i \times \{h^i\}.$$  

The reverse implication in the definition of formulation is direct from the construction of $Q(\mathcal{P}, H)$. □

In the sequel we assume all families of polyhedra considered satisfy the equal recession cones condition.

**Assumption 1.** The family of polyhedra $\mathcal{P} := \{P^i\}_{i=1}^n$ is such that $P^i_{\infty} = P^j_{\infty}$ for all $i,j \in [n]$.

A key advantage of embedding formulations is their flexible use of encodings. However, the following lemma shows that different encodings may lead to the same embedding formulation.

**Lemma 1.** Let $\mathcal{P} := \{P^i\}_{i=1}^n$ be a finite family of non-empty polyhedra in $\mathbb{R}^d$, $H \in \mathcal{H}_{k_1}(n)$ and $G \in \mathcal{H}_{k_2}(n)$. If there exists an affine map $A : \mathbb{R}^{k_1} \to \mathbb{R}^{k_2}$ such that $A$ is a bijection between $\text{conv}(H)$ and $\text{conv}(G)$ then $Q(\mathcal{P}, H)$ is affinely isomorphic (and hence combinatorially equivalent) to $Q(\mathcal{P}, G)$.

In particular, for fixed $\mathcal{P} = \{P^1, P^2\}$, all $Q(\mathcal{P}, H)$ for $H \in \mathcal{H}(2)$ are affinely isomorphic.

**Proof.** The first part of the statement is straightforward. For the second statement, given $H = \{h^1, h^2\}$, $G = \{g^1, g^2\} \in \mathcal{H}(2)$ take

$$A(h) = (g^2 - g^1) \frac{(h^2 - h^1) \cdot h}{(h^2 - h^1) \cdot (h^2 - h^1)} + g^1 \frac{(h^2 - h^1) \cdot h}{(h^2 - h^1) \cdot (h^2 - h^1)}.$$

□

A direct corollary of Lemma 1 is that all variants of a unary encoded embedding formulation (i.e. those obtained by changing the order in $\mathcal{P}$) lead to equivalent formulations, which justifies our unique unary encoding statement. In contrast, as we will see in Section 5, the equivalence may not hold among binary encodings. Finally, we note that the unary encoding is equivalent to an encoding used in commonly denoted incremental formulations (e.g. [50, Section 8] and [58]).

Both the size of an arbitrary formulation and the size of an embedding formulation lead to a natural complexity measure for unions of polyhedra.
Definition 6 (Relaxation and Embedding Complexity). For a polyhedron $Q$ let $\text{size}(Q)$ be equal to the number of facet defining inequalities of $Q$. Then, for a family of polyhedra $\mathcal{P} := \{P^n\}_{i=1}^n$ we let its relaxation complexity be
\[
\text{rc}(\mathcal{P}) := \min \left\{ \text{size}(Q) : Q \text{ is a polyhedron, } H \in \mathcal{H}(n) \text{ and } (Q, H) \text{ is a formulation for } \bigcup_{i=1}^n P^n \right\}
\]
and its embedding complexity be
\[
\text{mc}(\mathcal{P}) := \min \{ \text{size}(Q(\mathcal{P}, H)) : H \in \mathcal{H}(n) \}.
\]

2.1. Relation to other complexity measures The relaxation complexity of a family of polytopes is a direct analog of the relaxation complexity of a set of integer points introduced in [31]. However, the natural analog of the embedding complexity for a set of integer points is not so clear. One possible analog is to take the size of the convex hull of the set points, as this yields the smallest ideal (or integral) IP formulation of the set without auxiliary variables. Another possible analog is to consider the extension complexity of the convex hull of the set points, that is the smallest extended formulation of this convex hull (e.g. see [12, 29] and the references therein), which yields the smallest ideal IP formulation of the set with auxiliary variables. Both these notions have the following natural adaptation to unions of polyhedra.

Definition 7 (Hull and Extension Complexity). For a family of polyhedra $\mathcal{P} := \{P^n\}_{i=1}^n$ we let its hull complexity be
\[
\text{hc}(\mathcal{P}) := \text{size} \left( \text{conv} \left( \bigcup_{i=1}^n P^n \right) \right)
\]
and its extension complexity
\[
\text{xc}(\mathcal{P}) := \min \left\{ \text{size}(R) : \text{proj}_{x,y}(R) = \text{conv} \left( \bigcup_{i=1}^n P^n \right) \right\}.
\]

Unfortunately, these adaptations are not formulation complexities as neither $\text{conv} \left( \bigcup_{i=1}^n P^n \right)$ nor the polyhedron $R$ that achieves the minimum for the extension complexity yield a MIP formulation for $\bigcup_{i=1}^n P^n$. A simple fix for this issue is to combine the notions of embedding and extension complexities as follows.

Definition 8 (Extended Relaxation and Embedding Complexity). Let $\mathcal{P} := \{P^n\}_{i=1}^n$ be a finite family of polyhedra in $\mathbb{R}^d$, $H \in \mathcal{H}(n)$ and $R \subseteq \mathbb{R}^{d+k+r}$ be a polyhedron. We say $(R, H)$ is an extended formulation of $\bigcup_{i=1}^n P^n$ if and only if $(\text{proj}_{x,y}(R), H)$ is a formulation of $\bigcup_{i=1}^n P^n$ and an extended embedding formulation\(^1\) if and only if
\[
\text{proj}_{x,y}(R) = Q(\mathcal{P}, H).
\]

We let the extended relaxation complexity of $\mathcal{P}$ be
\[
\text{xrc}(\mathcal{P}) := \min \left\{ \text{size}(R) : \text{R is a polyhedron, } H \in \mathcal{H}(n) \text{ and } (\text{proj}_{x,y}(R), H) \text{ is a formulation for } \bigcup_{i=1}^n P^n \right\},
\]

\(^1\)Here we follow the usual convention of not including the number of equations or variables (e.g. [29]). We will refine this notion of size for a special class of polyhedra in Section 3.

\(^2\)By an analog of Proposition 1, extended embedding formulations are essentially equivalent to extended ideal mixed 0-1 formulations. We utilize the former nomenclature to emphasize the connection to the central topic of this paper.
and its extended embedding complexity be
\[
\text{xmc}({\mathcal P}) := \min \left\{ \text{size}(R) : R \text{ is a polyhedron, } H \in \mathcal{H}(n) \text{ and } \text{proj}_{x,y}(R) = Q({\mathcal P}, H) \right\}.
\]

Non-extended formulations and formulation complexities only allow the 0-1 variables required to construct a formulation as auxiliary variables. In contrast, extended formulations and formulation complexities allow any kind of auxiliary variables (required 0-1 variables plus other continuous or integer constrained variables). These additional variables greatly strengthen and simplify the formulations. For instance, using a standard formulation by Balas, Jeroslow and Lowe [6, 28, 35] we directly obtain a linear upper bound for both extended complexities.

**Proposition 2.** For a family of polyhedra \( \mathcal{P} := \{P^i\}_{i=1}^n \) there exists an ideal MIP formulation of \( \bigcup_{i=1}^n P^i \) whose LP relaxation is a polyhedron \( R \subseteq \mathbb{R}^{n \times (d+1)} \) with size\( (R) = n + \sum_{i=1}^n \text{size} (P^i) \). Hence,
\[
\text{xrc}({\mathcal P}) \leq \text{xmc}({\mathcal P}) \leq n + \sum_{i=1}^n \text{size} (P^i).
\]

**Proof.** The Balas, Jeroslow and Lowe formulation can be found in Proposition 4.2 of [50]. □

In contrast to the linear upper bound from Proposition 2, it is easy to construct examples where \( \text{hc}({\mathcal P}) \) is exponential on the natural size of \( \mathcal{P} \) and Example 2 in subsection 2.2 shows the same for \( \text{mc}({\mathcal P}) \). This suggests that the auxiliary variable restrictions of the non-extended formulations may be unnecessary and costly. However, we will see that careful selection of the encoding yield non-extended formulations with sizes smaller than the bounds from both Proposition 2 and an improved bound we describe in Proposition 4. Furthermore, the associated non-extended embedding formulations have been shown to computationally outperform the extended formulations from both these propositions [51, 52].

We end this subsection with the following straightforward lemma that describes simple relations between the complexity measures.

**Lemma 2.** For a family of polyhedra \( \mathcal{P} := \{P^i\}_{i=1}^n \) we have
- \( \text{xc}({\mathcal P}) \leq \text{hc}({\mathcal P}) \),
- \( \text{xc}({\mathcal P}) \leq \text{xmc}({\mathcal P}) \leq \text{mc}({\mathcal P}) \),
- \( \text{xrc}({\mathcal P}) \leq \text{xmc}({\mathcal P}) \),
- \( \text{xrc}({\mathcal P}) \leq \text{rc}({\mathcal P}) \), and
- \( \text{rc}({\mathcal P}) \leq \text{mc}({\mathcal P}) \).

An interesting relation that is missing from Lemma 2 is that between the embedding and hull complexities. We now explore the relation between these complexities and the Minkowski sum of the family. This discussion serves as one motivation to the use of non-unary encodings as a path to construct embedding formulations with smaller sizes than the extended embedding formulation from Corollary 2. We further study the relation between complexity measures in Section 7.

### 2.2. Embedding formulations, convex hull, minkowski sums and projections

Let \( \mathcal{P} := \{P^i\}_{i=1}^n \). Then a direct property of \( Q({\mathcal P}, H) \) is that its projection onto the \( x \) variables is equal to \( \text{conv} (\bigcup_{i=1}^n P^i) \). However, because the projection operation can significantly increase the number of inequalities of a polyhedron, the size of \( \text{conv} (\bigcup_{i=1}^n P^i) \) does not yield a direct lower bound on the size of \( Q({\mathcal P}, H) \) (unless the dimension of the encoding is very small). On the other hand, \( Q({\mathcal P}, H) \) for \( H = \{e^i\}_{i=1}^n \) includes the Minkowski sum \( P^1 + P^2 + \ldots + P^n := \{\sum_{i=1}^n x^i : x^i \in P^i \text{ } \forall i \in [n]\} \) through Cayley trick. More precisely, for \( H = \{e^i\}_{i=1}^n \) we have that \( \{x \in \mathbb{R}^n : (x,1/n)\sum_{i=1}^n e^i \in Q({\mathcal P}, H)\} \) is combinatorially equivalent to \( P^1 + P^2 + \ldots + P^n \) and hence we get the following lemma.
Lemma 3. Let $\mathcal{P} := \{P_i\}_{i=1}^{n}$ be a family of polyhedra in $\mathbb{R}^d$. If $H$ is the unary encoding, then

$$\text{size}(P^1 + P^2 + \ldots + P^n) \leq \text{size}(Q(\mathcal{P}, H)).$$

Now, it is easy to construct examples where the size of $\text{conv}(\bigcup_{i=1}^{n} P_i)$ is larger than the size of $P^1 + P^2 + \ldots + P^n$ (e.g. take $\mathcal{P} = \{\text{conv} (\{-e^i, e^i\})\}_{i=1}^{n}$). However, using results from [9, 24] we can also construct an example where the Minkowski sum is significantly larger than the convex hull of the union.

Example 2. Consider the dual pair of simplices $P^1 = \text{conv} (\{v^i\}_{i=0}^{n})$ and $P^2 = \text{conv} (\{w^i\}_{i=0}^{n})$ where $\{v^i, w^i\}_{i=0}^{n} \subseteq \mathbb{R}^n$ are given by

$$v^j_i = \begin{cases} n & j = i \\ -1 & j \neq i \end{cases}, \quad w^j_i = \begin{cases} -1 & j = i \\ 0 & j \neq i \end{cases}$$

for all $j \in \{1, \ldots, n\}$ and $v^0_j = -w^0_j = -1$ for all $j \in \{1, \ldots, n\}$. By Theorem 6 in [24] we have that the size of $P^1 + P^2$ is $2^{n+1} - 2$. However, it is easy to see that $\text{hc}(\mathcal{P})$ is $O(n^2)$. Furthermore, by Lemma 1, all $Q(\mathcal{P}, H)$ are combinatorially equivalent to the polyhedron obtained for $H = \{0, 1\}$. For this last choice of $H$ we have that $Q(\mathcal{P}, H)$ is the antiprism of $P^1$ whose size is $2^{n+1}$ [9], and hence $\text{me}(\mathcal{P}) = 2^{n+1}$.

Example 2 is strongly related to Example 4, Lemma 4.1 and Example 5 in [50] (e.g. the first part of Lemma 4.1 can be obtained as a direct corollary of Theorem 6 in [24]). The examples from [50] show that both rc($\mathcal{P}$) and xmc($\mathcal{P}$) are $O(n)$, which is in stark contrast with the exponential embedding complexity. However, Example 2 has two somewhat atypical characteristics that restrict the power of embedding formulations. The first one is that the example does not allow for any flexibility on the encoding selection; by Lemma 1 $Q(\mathcal{P}, \{0, 1\})$ is effectively the only option for the union of two polyhedra. The second one is that, as noted in the second part of Lemma 4.1 from [50], constructing a polynomial sized ideal formulation of $P^1 \cup P^2$ requires a number of auxiliary variables (binary or continuous) that grows with $n$ and, because of the unique encoding option, $Q(\mathcal{P}, H)$ effectively only uses one. In Sections 5 and 6 we will see that when these two issues are avoided we sometimes can construct embedding formulations that are comparable or smaller than the best known alternative formulation. One key to achieving this is to have a growing number of polyhedra so the dimension of $H$ can grow (i.e. so the number of binary variables grows). Another is to use an encoding which prevents the Cayley trick from including the Minkowski sum in $Q(\mathcal{P}, H)$.

One natural way to achieve the former is to use a binary encoding $H \subseteq \{0, 1\}^{\lceil \log_2 n \rceil}$ as the associated $Q(\mathcal{P}, H)$ only has potential to contain partial sums of $O(\log_2 n)$ of the $n$ polyhedra. However, the following simple lemma, whose proof is analogous to Lemma 4.1 from [50], shows that the binary encoding can be at a disadvantage when $\text{conv}(\bigcup_{i=1}^{n} P_i)$ has a large size.

Lemma 4. Let $\mathcal{P} := \{P_i\}_{i=1}^{n}$ be a finite family of non-empty polyhedra in $\mathbb{R}^d$, $H := \{h^i\}_{i=1}^{n} \subseteq \{0, 1\}^k$ be an encoding of $\{n\}$ and $Q = Q(\mathcal{P}, H)$. Then

$$\text{size} \left( \text{conv} \left( \bigcup_{i=1}^{n} P_i \right) \right) = \text{size} \left( \text{proj}_x (Q) \right) \leq \left( \frac{\text{size} (Q)}{\text{size} (Q) - k - 1} \right).$$

Lemma 4 shows that the size of the projection of an embedding formulation with a unary encoding could be exponential on the size of the formulation, but the size of the projection for a binary encoded formulation can be at most quasipolynomial on the size of the formulation. This could place binary encoded formulations at a small disadvantage compared to unary encoded formulations when the size of $\text{conv}(\bigcup_{i=1}^{n} P_i)$ is large. However, in the following sections we consider an important class of disjunctive constraints for which $\text{conv}(\bigcup_{i=1}^{n} P_i)$ is trivial and where the careful selection of a binary encoding can yield significant improvements in formulation complexity.
3. **V-formulations** To simplify our analysis of formulation complexities we now consider a class of disjunctions composed of unions of specially structured faces of a common simplex. These faces are obtained from the simplex by fixing certain variables to zero and can be defined as follows.

**Definition 9.** Let $V$ be a finite set, $\Delta^V := \{ \lambda \in \mathbb{R}^V : \sum_{v \in V} \lambda_v = 1 \}$ be the standard simplex in $\mathbb{R}^V$ and $\mathcal{T} := \{ T_i \}_{i=1}^n$ be a family of subsets of $V$. We let $\mathcal{P}(\mathcal{T}) := \{ P(T_i) \}_{i=1}^n$ where

$$P(T_i) := \{ \lambda \in \Delta^V : \lambda_v \leq 0 \quad \forall v \notin T_i \}$$

for each $i \in [n]$. In addition, for any $H := \{ h^i \}_{i=1}^n \in \mathcal{H}_k(n)$ we let

$$Q(\mathcal{T}, H) := Q(\mathcal{P}(\mathcal{T}), H) = \text{conv} \left( \{ (\lambda, y) \in \mathbb{R}^V \times \mathbb{R}^k : \exists i \in [n] \text{ s.t. } \lambda \in P_i, \quad y = h^i \} \right).$$

Finally, we let $\text{rc}(\mathcal{T}) := \text{rc}(\mathcal{P}(\mathcal{T}))$ and $\text{mc}(\mathcal{T}) := \text{mc}(\mathcal{P}(\mathcal{T}))$.

As described in the following definition, constraints of this form include the special ordered sets of type 1 (SOS1) and the special ordered sets of type 2 (SOS2) introduced by Beale and Tomlin [7].

**Definition 10.** We say $\mathcal{T} := \{ T_i \}_{i=1}^n$ is an SOS1 constraint on $V := [n]$ if and only if $T_i = \{ i \}$ for $i \in [n]$. We also say $\mathcal{T} := \{ T_i \}_{i=1}^n$ is an SOS2 constraint on $V := [n+1]$ if and only if $T_i = \{ i, i+1 \}$ for $i \in [n]$.

Furthermore, specially structured unions of polyhedra from Definition 9 are strongly related to commonly denoted $\mathcal{V}$-formulations of unions of arbitrary polyhedra (see Corollary 5.2 and 6.3 in [50]). In fact, the following straightforward proposition shows that arbitrary unions of polyhedra are essentially linear images of these specially structured polyhedra.

**Proposition 3.** Let $\mathcal{R} := \{ R_i \}_{i=1}^n \subseteq \mathbb{R}^d$ be a finite family of non-empty polyhedra with identical recession cones ($R_{i\infty} = R_{i\infty}$ for all $i, j \in [n]$),

$$V := \bigcup_{i=1}^n \text{ext}(R_i)$$

and $\mathcal{T} := \{ T_i \}_{i=1}^n$ with $T_i = \text{ext}(R_i)$. Then $x \in \bigcup_{i=1}^n R_i$ is equivalent to

$$\sum_{i=1}^n \sum_{v \in T_i} v \lambda_v + \sum_{r \in \text{ray}(R_i)} r \mu_r = x \quad (8a) \quad \text{with} \quad \lambda \in \bigcup_{i=1}^n P(T_i) \quad (8b) \quad \mu \in \mathbb{R}^{\text{ray}(R_i)}. \quad (8c)$$

Hence, if $(Q, H)$ is a formulation of $\bigcup_{i=1}^n P(T_i)$ for some $H \in \mathcal{H}(m)$, then a formulation for $\bigcup_{i=1}^n R_i$ is obtained by replacing (8b) with $(\lambda, y) \in Q$ and $y \in \mathbb{Z}^k$. In particular, if $Q = Q(\mathcal{T}, H)$, then the resulting formulation is ideal.

The relation with $\mathcal{V}$-formulations also allows for a slight improvement of Proposition 2.

**Proposition 4.** For $\mathcal{T} := \{ T_i \}_{i=1}^n$ there exist an ideal MIP formulation of $\bigcup_{i=1}^n P(T_i)$ whose Linear Programming relaxation is a polyhedron $R \subseteq \mathbb{R}^{\sum_{i=1}^n |T_i| + \lceil \log_2 n \rceil}$ with size$(R) = \sum_{i=1}^n |T_i|$. Hence,

$$\text{xrc}(\mathcal{T}) \leq \text{xmc}(\mathcal{T}) \leq \sum_{i=1}^n |T_i|.$$
Proof. The formulation is a natural adaptation of the formulation in Proposition 9.3 of [50] and is given by

\[
\sum_{i \in [n]: v \in T_i} \lambda^i_v = \lambda_v \quad \forall v \in V \\
\sum_{i=1}^n \sum_{v \in T_i} \lambda^i_v = 1 \\
\sum_{i=1}^n \sum_{v \in T_i} h^i \lambda^i_v = y \\
\lambda^i_v \geq 0 \\
y \in \{0, 1\}^{\lceil \log_2 n \rceil},
\]

where \( H := \{h^i\}_{i=1}^n \in H_{\lceil \log_2 n \rceil}(n) \). \( \square \)

In the sequel we assume the following simplifying assumption that ensures no \( \lambda \) variable is always fixed to zero.

**Assumption 2.** \( \bigcup_{i=1}^n T_i = V \).

A convenient property of families of polyhedra from Definition 9 is that they are 0-1 polytopes with purely combinatorial structure and no numerical data. Indeed, when considering arbitrary unions of polyhedra through Proposition 3 all numerical data associated to the union is confined to (8a) and (8b) is restricted to the combinatorial structure induced by possible common vertices of the original polyhedra. This purely combinatorial structure greatly simplifies the analysis of the formulation complexities associated to these unions of polyhedra. For instance, because \( P(T_i) \) and \( Q(T, H) \) are 0-1 polytopes we have that variable bounds are geometrically special as they are facets of the underlying unit cube. They are also special from a practical computational perspective when the LP relaxation of the formulations is solved by the simplex algorithm. For this reason we separate variable bounds from other inequalities in our size accounting. This separation also forces us to explicitly account for equalities. Indeed, the reason we did not previously consider them is that they can be removed by eliminating variables. However, this elimination could transform bound inequalities into general inequalities confounding our desired separation. We can check that the remaining potential inequalities in a minimal description of \( Q(T, H) \) could include inequalities in a minimal description of \( \text{conv}(H) \) and inequalities that consider both the \( \lambda \) and \( y \) variables. This last set of inequalities is precisely the one associated to Minkowski sums in the Cayley trick and compose the most critical structure of an embedding formulation, so we denote them embedding inequalities.

**Definition 11.** Let

\[
\sum_{v \in V} a_v \lambda_v + \sum_{i=1}^n b_i y_i = a \cdot \lambda + b \cdot y \leq c \tag{9}
\]

be a generic inequality for a formulation \( Q \) of \( \bigcup_{i=1}^n P(T_i) \). We say (9) is an embedding inequality if and only if \( \sum_{v \in V} |\lambda_v| > 0 \) and \( \sum_{i=1}^n |b_i| > 0 \).

We let \( \text{size}_B(Q) \), \( \text{size}_E(Q) \) and \( \text{size}_M(Q) \) the restriction of \( \text{size}(Q) \) to variable bounds, embedding inequalities and equations respectively. We similarly define \( \text{rc}_B(T) \), \( \text{rc}_E(T) \), \( \text{rc}_M(T) \), \( \text{mc}_B(T) \), \( \text{mc}_E(T) \) and \( \text{mc}_M(T) \).

In the rest of this section we first consider a simple case where the embedding complexity can be determined exactly and then consider basic properties of embedding formulations for more interesting structures.

\*\*\* We count embedding inequalities that are implied equations twice. However, in Section 3.2 we will see that under mild conditions there are no embedding inequalities that are implied equations. \*\*\*
3.1. Optimal embedding formulation for disjoint case  When the elements of $T$ are disjoint we can obtain a description of $Q(T,H)$ as a direct corollary of Proposition 9.3 in [50] and Proposition 1.

Proposition 5.  Let $H := \{h^i\}_{i=1}^n \in \mathcal{H}_k(n)$ and $T := \{T_i\}_{i=1}^n$ be such that $T_i \cap T_j = \emptyset$ for all $i \neq j$. Then $Q(T,H)$ can be described by

\[
\begin{align*}
\sum_{i=1}^n \sum_{v \in T_i} h^i \lambda_v &= y & \quad (10a) \\
\sum_{v \in V} \lambda_v &= 1 & \quad (10b) \\
\lambda &\in \mathbb{R}^V_+ & \quad (10c) \\
y &\in [0,1]^k. & \quad (10d)
\end{align*}
\]

A simple analysis of this description then yields the exact embedding complexity for this disjoint case.

Proposition 6.  $T_i \cap T_j = \emptyset$ for all $i \neq j$ then

- $mc(T) = mc_B(T) = n,$
- $mc_M(T) = 2 \lceil \log_2 n \rceil$ and,
- $mc_E(T) = \lceil \log_2 n \rceil + 1.$

Proof.  We can check that constraints $\lambda_v \geq 0$ for $v \in V$ are facet defining for (10) while constraints $0 \leq y_i \leq 1$ for $i \in [k]$ are redundant for (10). In addition, if any of the equalities in (10) is linearly dependent on the others we can eliminate one of the equalities that includes $y$ and reduce the dimension of $y$. The result follows by noting that the smallest $k$ that for which exists $H \subseteq \{0,1\}^k$ such that $H \in \mathcal{H}(n)$ is $k = \lceil \log_2 n \rceil$. □

It is interesting to note that only a logarithmic number of embedding inequalities are needed in this case, but in Section 4 we will show that if we are willing to forgo idealness of the formulation only a constant (independent of $n$) number of embedding inequalities are enough.

3.2. Basic properties for connected case  To analyze the embedding complexity of more complicated cases it will be convenient to prove some basic properties of embedding formulations. A natural simplifying assumption for the non-disjoint case is the following connectivity condition.

Assumption 3 (Connectedness).  For all $j \neq j'$ there exist $\{i_1,\ldots,i_r\} \subseteq [n]$ such that $j = i_1$, $j' = i_r$ and $T_{i_l} \cap T_{i_{l+1}} \neq \emptyset$ for all $l \in [r-1]$. 

We note that many of the properties we consider extend to the non-connected case by considering each connected components separately (e.g. see [33] for an example with the unary encoding). However, for simplicity and briefness we now assume the connectivity requirement except when considering the disjoint case.

Another simplifying assumption we often use is sharpness and encoding sharpness of the formulation. The reason for this is that valid formulations that do not satisfy these conditions may not imply the equalities of $(Q,H)$ or the inequalities describing $\text{conv}(H)$ (e.g. see Example 8 in the appendix).

We begin by analyzing the implied equalities of a sharp formulation, for which we require the following definition.

Definition 12.  For $H := \{h^i\}_{i=1}^n$, let $L(H) := \text{aff}(H) - h^1$ be the linear space parallel to the affine hull of $H$.

The following lemma is a direct generalization of Proposition 1 from [33] whose proof we include for completeness.
Lemma 5. Let \((Q, H)\) be a sharp and encoding-sharp formulation. If Assumption 3 is satisfied, then \(\dim(Q) = |V| + \dim(H) - 1\) and the constraints defining \(\text{aff}(Q)\) are precisely
\[
\sum_{v \in V} \lambda_v = 1 \tag{11}
\]
and the constraints defining \(\text{aff}(H)\) (i.e. equations that do not include the \(\lambda\) variables)

Proof. First note that because of the two sharpness conditions, both (11) and the constraints defining \(\text{aff}(H)\) must be valid for \(Q\). Then, \(\dim(Q) \leq |V| + \dim(H) - 1\) follows directly by noting that (11) is linearly independent from any equality that only consider the \(y\) variables and that any equality that only considers the \(y\) variables is implied by the equalities defining \(\text{aff}(H)\).

For the reverse inequality let
\[
a \cdot \lambda + b \cdot y = c \tag{12}
\]
be an arbitrary equality of \(Q\). Suppose for a contradiction that (12) is linearly independent to (11) and the equations defining \(\text{aff}(H)\). By subtracting multiples of (11) and the equations defining \(\text{aff}(H)\) we may assume \(a = 0\) and \(b \in L(H)\). Letting \((\lambda, y) = (e^v, h^i)\) for \(v \in T_i\) in (12) we have that \(b \cdot h^i = -a_v e^v\). Then, by Assumption 3 we have that \(b \cdot h^i = -a_v e^v\) for all \(i, j \in [n]\) and \(a_u = a_v\) for all \(u, v \in V\). This implies \(b \cdot (h^i - h^j) = 0\) for all \(i \in [n]\), which together with \(b \in L(H)\) implies \(b = 0\) and \(a = 0\). This contradicts our assumption of the linear independence of (12). □

Next we analyze the necessity of inequalities \(\lambda_v \geq 0\), for which we require the following definition.

Definition 13. We say \(v \in V\) is a cut vertex of \(\mathcal{T}\) if \(\mathcal{T}' := \{T_i \setminus \{v\}\}_{i=1}^m\) does not satisfy connectedness assumption 3.

The following lemma is a direct generalization of one direction of Proposition 7 from [33]. We again include the proof for completeness.

Lemma 6. If \(Q\) is a sharp formulation, Assumption 3 is satisfied, \(|T_i| \geq 2\) for all \(i \in [n]\) such that \(v \in T_i\), and \(v \in V\) is not a cut vertex of \(\mathcal{T}\), then \(\lambda_v \geq 0\) describes a facet of \(Q\).

Proof. If \(Q\) is sharp, then inequality \(\lambda_v \geq 0\) describes a face of \(Q\) as it is a valid inequality for it and is satisfied at equality for all points \((e^u, h^i)\) for \(u \in T_i \setminus \{v\}\). Let
\[
a \cdot \lambda + b \cdot y \leq c \tag{13}
\]
be a valid inequality of \(Q\) that induces a facet containing the face induced by \(\lambda_v \geq 0\). Without loss of generality we may again assume \(c = 0\) and \(b \in L(H)\). Because \(v\) is not a cut vertex, (13) is satisfied at equality by \((e^u, h^i)\) for \(u \in T_i \setminus \{v\}, T_i \setminus \{v\} \neq \emptyset\) for all \(i \in [n]\) and Assumption 3, we have that \(b \cdot h^i = b \cdot h^j\) for all \(i, j \in [n]\) and \(a_u = a_v\) for all \(u, u' \in V \setminus \{v\}\). This implies \(b \cdot (h^i - h^j) = 0\) for all \(i \in [n]\), which together with \(b \in L(H)\) implies \(b = 0\) and \(a_u = 0\) for all \(u \in V \setminus \{v\}\). Validity of (13) implies \(a_v \leq 0\) and being facet defining further implies \(a_v < 0\). Then (13) is a positive multiple of \(\lambda_v \geq 0\) and hence \(\lambda_v \geq 0\) is facet defining. □

For an case where condition \(|T_i| \geq 2\) does not hold and there exists \(v \in V\), which is not a cut vertex, but \(\lambda_v \geq 0\) is not facet defining see Example 9 in the appendix. Further analysis of when \(\lambda_v \geq 0\) is facet defining seems strongly dependent on the encoding used. For instance for the unary encoding we have the following result from [33].

Lemma 7 ([33]). If \(H\) is the unary encoding, Assumption 3 is satisfied and \(v \in V\) is a cut vertex of \(\mathcal{T}\), then \(\lambda_v \geq 0\) is not facet defining for \(Q(\mathcal{T}, H)\).

In contrast, the following example shows that determining when \(\lambda_v \geq 0\) is not facet defining is less clear for binary encodings.

Example 3. Let \(\mathcal{T}\) be an SOS2 constraint on \(V = [5]\) and \(H = \{(0,1)^T, (1,1)^T, (1,0)^T, (0,0)^T\}\). Then \(\lambda_v \geq 0\) is facet defining for \(Q(\mathcal{T}, H)\) for cut vertex \(v = 2\), but not for cut vertex \(v = 3\)
Our final results studies when inequalities for \( \text{conv} (H) \) are facet defining for \( Q (T, H) \).

**Proposition 7.** Let \( T \) be such that there is no \( i, j \in [n] \), \( i \neq j \) such that \( T_i \subseteq T_j \). Then none of the facet defining inequalities of \( \text{conv} (H) \) is facet defining for \( Q (T, H) \).

**Proof.** Consider an inequality
\[
b \cdot y \leq c
\]
that is facet defining for \( \text{conv} (H) \) and assume that it is facet defining for \( Q (T, H) \). We will reach a contradiction by constructing a face of \( Q (T, H) \) that strictly contains the face induced by (14).

To achieve this we will use the so-called *facet procedure* [3, 11, 20].

First note that (14) cannot be facet defining for \( Q (T, H) \) unless
\[
a_v := \min \{ c - b \cdot h^i : i \in [n], \ v \in T_i \} = 0
\]
for all \( v \in V \) (if \( a_v < 0 \) the inequality is invalid and if \( a_v > 0 \) it can be strengthened by changing the coefficient of \( \lambda_v \) from 0 to \( a_v \)). Now, let \( I := \{ i \in [n] : b \cdot h^i = c \} \) and \( i_0 \in [n] \setminus I \). Because, \( c - b \cdot h^{i_0} > 0 \) and \( a_v = 0 \) for all \( v \in T_{i_0} \) we have that \( T_{i_0} \subseteq \bigcup_{i \in I} T_i \). Then, by the assumption of non-containment between sets in \( T \) we must have \( v_1, v_2 \in T_{i_0} \) such that \( v_1 \neq v_2 \). We have that \( h^{i_0} \) is affinely independent of \( \{ h^i \}_{i \in I} \) (the later satisfy (14) at equality and the later satisfies it strictly) and hence \((h^{i_0}, e^{v_2})\) is affinely independent of \( \{(h^i, e^v) : i \in I, v \in T_i \}\). Then there exist \( \bar{c} \in \mathbb{R}, \bar{b} \in \mathbb{R}^k \) and \( \bar{a} \in \mathbb{R}^v \) such that
\[
\bar{c} - \bar{b} \cdot h^i - \bar{a} \cdot e^v = 0 \quad \forall i \in I, \ v \in T_i
\]
(15)
\[
\bar{c} - \bar{b} \cdot h^{i_0} - \bar{a} \cdot e^{v_1} = 0
\]
(16)
\[
\bar{c} - \bar{b} \cdot h^{i_0} - \bar{a} \cdot e^{v_2} > 0.
\]
(17)

Let \( A = \{(i, v) \in [n] \times V : i \in [n], \ v \in T_i, \ c - b \cdot h^i - a \cdot e^v < 0\} \). If \( A = \emptyset \) then \( \bar{a} \cdot \lambda + \bar{b} \cdot y \leq \bar{c} \) is valid for \( Q (T, H) \) and it defines a proper face \((h^{i_0}, e^{v_2})\) satisfies it strictly). Furthermore, the face it defines strictly contains the face defined by (14) (it contains the extra point \((h^{i_0}, e^{v_1})\)). This contradicts (14) being a facet. If \( A \neq \emptyset \) let
\[
\alpha := \max_{(i, v) \in A} \frac{\bar{a} \cdot e^v + \bar{b} \cdot h^i - \bar{c}}{c - b \cdot h^i} > 0
\]
and let \((i^*, v^*)\) one element in \( A \) than achieves this maximum. Then \( \bar{a} \cdot \lambda + (\bar{b} + \bar{a} \cdot \alpha) \cdot y \leq \bar{c} + \alpha c \) is valid for \( Q (T, H) \) and it defines a proper face \((h^{i_0}, e^{v_2})\) satisfies it strictly). Furthermore, the face it defines strictly contains the face defined by (14) (it contains the extra point \((h^{i_0}, e^{v_1})\)). This again contradicts (14) being a facet. \( \square \)

For a case where the containment condition fails and a facet defining inequality of \( \text{conv} (H) \) is also facet defining for \( Q (T, H) \) see Example 9 in the appendix.

**4. Bounds on relaxation complexity of \( V \)-formulations** We now study the relaxation complexity of unions of polyhedra associated to \( V \)-formulations. We begin by considering lower bounds on the relaxation complexity of these unions for which we need the following definition from extremal set theory (e.g. [19]).

**Definition 14.** A family \( \{ A_i \}_{i \in I} \subseteq 2^L \) is called *r-cover-free* if and only if
\[
A_i \setminus \bigcup_{j \in J} A_j \neq \emptyset \quad \forall i \in I, \text{ and } J \subseteq I \setminus \{ i \} \text{ with } |J| = r.
\]
(18)

We let \( f_r (L) \) denote the maximum cardinality of a \( r \)-cover-free family in \( 2^L \).
Through this definition we obtain the following lower bounds on the number of embedding inequalities of any formulation for union of polyhedra associated to $\mathcal{V}$-formulations.

**Proposition 8.** Let $V$ be a finite set and $\mathcal{T} := \{T_i\}_{i=1}^n$ be a family of subsets of $V$.
1. If $n \geq 2$, then $\text{rc}_M(\mathcal{T}) \geq 2$.
2. If $I \subseteq \{n\}$ is such that $T_i \cap T_j \neq \emptyset$ for all $i, j \in I$, then
   \[
   \left( \frac{\text{rc}_M(\mathcal{T})}{\lfloor \text{rc}_M(\mathcal{T})/2 \rfloor} \right) \geq |I|.
   \]
   In particular,
   • $\text{rc}_M(\mathcal{T}) \geq 2.8 \ln(|I|)$ for $|I| \geq 5$.
3. If $I \subseteq \{n\}$ and $r \in [2, n - 1]$ is such that $\bigcap_{j \in I} T_j \setminus T_i \neq \emptyset$ for all $i \in I$ and $J \subseteq \{n\} \setminus \{i\}$ with $|J| = r$, then $|I| \leq f_r(\text{rc}_M(\mathcal{T}))$. In particular,
   • if the condition holds for $r = 2$, then $\text{rc}_M(\mathcal{T}) \geq 4.48 \ln(|I|)$, and
   • if the condition holds for $r$ such that $|I| < \left\lceil \frac{r+2}{2} \right\rceil$, then $\text{rc}_M(\mathcal{T}) \geq |I|$.

**Proof.** Consider a formulation of $\bigcup_{i=1}^n P(T_i)$ given by
   \[
   \sum_{v \in V} a_{vl}^i \lambda_v \leq c_l^i - \sum_{j=1}^k b_{lj}^i y_j \quad l \in [L],
   \]
   \[
   \sum_{v \in V} \lambda_v = 1
   \]
   \[
   -\lambda_v \leq 0 \quad \forall v \in V
   \]
   \[
   0 \leq \bar{c}^r - \sum_{j=1}^k \bar{b}_{lj}^r y_j \quad r \in [R],
   \]
   \[
   y \in \{0, 1\}^k
   \]
   where (19a) corresponds to the embedding inequalities of the formulation and the formulation has been possibly strengthened to include equation (19b), bounds (19c) and the facets of $\text{conv}(H)$ in (19d).

Now, for every $i \in \{n\}$, (19) with $y = h^i$ must imply
   \[
   \sum_{v \in V \setminus T_i} \lambda_v \leq 0.
   \]
   as this is one of the inequalities describing $P(T_i)$. This implication is equivalent to the existence of $\eta^i \in \mathbb{R}_+^L$, $\theta^i \in \mathbb{R}$, $\mu^i \in \mathbb{R}_+^V$ and $\nu^i \in \mathbb{R}_+^R$ such that
   \[
   \sum_{i=1}^L \eta_{li}^i a_{vl}^i + \theta^i - \mu_{vl}^i = \begin{cases} 
   1 & v \in V \setminus T_i \\
   0 & v \in T_i
   \end{cases}
   \]
   \[
   \sum_{i=1}^L \eta_{hi}^i \left( c_l^i - \sum_{j=1}^k b_{lj}^i h_j^i \right) + \theta^i + \sum_{r=1}^R \nu_{rr}^i \left( \bar{c}^r - \sum_{j=1}^k \bar{b}_{lj}^r h_j^i \right) = 0.
   \]
   Eliminating $\theta^i$ from (21) we obtain
   \[
   \sum_{i=1}^L \eta_{li}^i \left( a_{vl}^i - c_l^i + \sum_{j=1}^k b_{lj}^i h_j^i \right) + \sum_{r=1}^R \nu_{rr}^i \left( \sum_{j=1}^k \bar{b}_{lj}^r h_j^i - \bar{c}^r \right) - \mu_{vl}^i = \begin{cases} 
   1 & v \in V \setminus T_i \\
   0 & v \in T_i
   \end{cases}
   \]
   We have that $y = h^i$ satisfies (19d), so $\sum_{j=1}^k \bar{b}_{lj}^r h_j^i - \bar{c}^r \leq 0$ and hence we may assume without loss of generality that $\nu_{rr}^i = 0$ for all $r \in [R]$ (we can reproduce the effect of $\nu_{rr}^i > 0$ with $\mu^i > 0$). Similarly,
for any \( v \in T_i \) we have that \( \lambda = e^v \) and \( y = h^i \) satisfies (19a) and hence \( a_i^v - c_i + \sum_{j=1}^k b_j^i h_j^i \leq 0 \). Then, if \( \eta^i_l > 0 \) we have that \( a_i^v = c_i - \sum_{j=1}^k b_j^i h_j^i \) for all \( v \in T_i \). In particular, we have that
\[
\forall l \in [L] \quad \eta^i_l > 0 \Rightarrow a_i^u = a_i^v \quad \forall u, v \in T_i. \tag{23}
\]
In addition, we may replace every occurrence of \( c_i - \sum_{j=1}^k b_j^i h_j^i \) in (22) with \( a_i^v \) for a fixed \( v_0 \in T_i \) to obtain
\[
\sum_{i=1}^I \eta^i_l (a_i^v - a_i^{v_0}) - \mu^i_v = 1
\]
for every \( v \in V \setminus T_i \). Hence, we must have that
\[
\forall v \in V \setminus T_i \quad \exists l \in [L] \text{ s.t. } \eta^i_l > 0 \wedge a_i^v > a_i^u \quad \forall u \in T_i. \tag{24}
\]
Let \( \text{supp}(\eta') := \{ l \in [L] : \eta^i_l > 0 \} \) be the support of \( \eta' \). Then (24) implies
\[
T_i \cap T_j = \emptyset \quad \wedge \quad \text{supp}(\eta') = \text{supp}(\eta') \quad \Rightarrow \quad |\text{supp}(\eta')| \geq 2,
\]
which shows the first result.
Combining (23) and (24) we have that for \( i, j \in [n] \) such that \( T_i \cap T_j \neq \emptyset \) we must have
\[
\text{supp}(\eta^i) \setminus \text{supp}(\eta^j) \neq \emptyset. \tag{25}
\]
If for all \( i, j \in I \) we have \( T_i \cap T_j \neq \emptyset \), then by (25) we have that \( \{ \text{supp}(\eta^i) \}_{i \in I} \) is an anti-chain in \( 2^{[L]} \) for containment. Then, by Sperner’s lemma (e.g. [17, Theorem 1.1.1]) we must have \( \binom{L}{L/2} \geq |I| \). Using standard bounds on the binomial coefficient, we obtain
\[
\frac{|L/2| \ln \binom{L}{L/2}}{\ln(|I|)} \geq 0,
\]
which implies the bound for \( |I| \geq 5 \).

Now, assume the conditions for the third part of the proposition hold. Under this condition (and by noting that it implies \( T_i \cap T_j \neq \emptyset \) for all \( i, j \in I \)) we may again combine (23) and (24) to strengthen (25) to \( \{ \text{supp}(\eta^i) \}_{i \in I} \) being an \( r \)-cover free family. Indeed, suppose for a contradiction that \( \text{supp}(\eta^i) \subseteq \bigcup_{j \in I} \text{supp}(\eta^j) \) for some \( i, j \) and let \( v \in \bigcap_{j \in J} T_j \setminus T_i \). By (24) there must exist \( l \in \text{supp}(\eta^j) \) such that \( a_i^v > a_i^u \) for all \( u \in T_i \). However, for any \( j \in J \) we also have \( l \in \text{supp}(\eta^j) \) and there exists \( v' \in T_i \cap T_j \). Hence, by (23) we have that \( a_i^u = a_i^{v'} \) for all \( u, v' \in T_i \cup T_j \), which results in a contradiction.

The specific bound for \( r = 2 \) is obtained by using Theorem 2 in [18] and the second specific bound comes from Proposition 3.4 in [19].

To show tightness of the linear lower bound in the third part of Proposition 8, we can use the following result, which follows easily from Proposition 1.

**Proposition 9.** Let \( T \) be the cyclic SOS(-1) constraints given by \( V = [n] \) and \( T_i = V \setminus \{i\} \) for all \( i \in [n] \). If \( H \) is the unary encoding, then \( Q(T, H) \) is equal to
\[
\lambda_i \leq 1 - y_i \quad \forall i \in [n] \quad \text{(26a)}
\]
\[
\sum_{i=1}^n \lambda_i = 1 \quad \text{(26b)}
\]

\(^1\)Here we are using the assumption \( \nu^i_r = 0 \) for all \( r \in [R] \). We are also abusing the valid replacement with \( a_i^{v_0} \) for \( \eta^i_l = 0 \), but this is inconsequential.
\[
\sum_{i=1}^{n} y_i = 1 \quad \text{(26c)}
\]
\[
\lambda_i \geq 0 \quad \forall i \in [n] \quad \text{(26d)}
\]
\[
y_i \geq 0 \quad \forall i \in [n] \quad \text{(26e)}
\]

Then, by noting that for cyclic SOS(-1) constraint the third condition in Proposition 8 holds with \( I = [n] \) and \( r = n - 1 \), we obtain the following corollary.

**Corollary 1.** If \( \mathcal{T} \) is a cyclic SOS(-1) constraint on \( V = [n] \) then
- \( \text{rc}_M(\mathcal{T}) = \text{mc}_M(\mathcal{T}) = n \), and
- \( n \leq \text{rc}(\mathcal{T}) \leq \text{mc}(\mathcal{T}) \leq 3n \).

Furthermore, the upper bound on \( \text{mc}(\mathcal{T}) \) is achieved by a formulation with only two equations.

Cyclic SOS(-1) constraints are a case where the relaxation and embedding complexities coincide and hence the smallest formulation is also ideal. In contrast, to show tightness of the constant lower bound in the first part of Proposition 8 requires a formulation that usually fails to be ideal. However, the formulation is sharp and has a constant number of embedding inequalities for a wide range of families of polyhedra.

**Proposition 10.** Let \( \{I_j\}_{j=1}^{p} \) be a partition of \([n]\) such that \( T_i \cap T_k = \emptyset \) for all \( i, k \in I_j, i \neq k \) and \( j \in [p] \). Also, for each \( j \in [p] \) and \( i \in I_j \cup \{0\} \) let \( \bar{a}^{i,j} := r_j - a^{i,j} \) where \( r_j > 0 \) and \( \{a^{i,j}\}_{i \in I_j \cup \{0\}} \) is a pairwise distinct family of vectors in \( S_+^2(r_j) := \{ a \in \mathbb{R}^2_+ : \|a\|_2 = r_j \} \) for all \( j \in [p] \). Then

\[
\sum_{i \in I_j} \sum_{v \in V \setminus I_j \cup I_j} \bar{a}^{i,j} \lambda_v + \sum_{v \in V \setminus I_j \cup I_j} \bar{a}^{0,j} \lambda_v \leq \sum_{i \in I_j} \bar{a}^{i,j} y_i + \left( \max_{i \in I_j \cup \{0\}} a^{i,j} \right) \sum_{i \in \bigcup_{k=1}^{p} I_k \setminus I_j} y_i, \quad \forall j \in [p] \quad \text{(27a)}
\]

\[
\sum_{v \in V \setminus I_j} \lambda_v = 1 \quad \text{(27b)}
\]

\[
\sum_{i=1}^{n} y_i = 1 \quad \text{(27c)}
\]

\[
\lambda \in \mathbb{R}^V_+, \quad y \in \{0,1\}^n \quad \text{(27d)}
\]

is a sharp (and encoding sharp) MIP formulation of \( \bigcup_{i=1}^{n} P(T_i) \).

**Proof.** Sharpness of the formulation is straightforward, so it suffices to show that \((\lambda, y)\) satisfies (27) if and only if there exists \( i \in [n] \) and \( u \in T_i \) such that \( \lambda = e^u \) and \( y = e^i \). The if part is straightforward. For the only if part we show that for \( y = e^i \) the (27a)-(27d) imply

\[
\sum_{v \in V \setminus T_i} \lambda_v \leq 0. \quad \text{(28)}
\]

For this let \( j \in [p] \) be such that \( i \in I_j \) and note that by construction of the \( \bar{a}^{i,j} \) there exist \( \eta' \in \mathbb{R}^2_+ \) such that \( a^{i,j}_1 \eta'_1 + a^{i,j}_2 \eta'_2 = 1 \) and \( a^{i,j}_1 \eta'_1 + a^{i,j}_2 \eta'_2 > 1 \) for all \( l \in (I_j \cup \{0\}) \setminus \{i\} \). Then, by Farkas' lemma, (27a) with \( y = e^i \) imply

\[
\sum_{v \in T_i} \lambda_v + \sum_{l \in [n] \setminus \{i\}} \sum_{v \in T_i} \bar{a}_l \lambda_v \leq 1, \quad \text{(29)}
\]

where \( \bar{a}_l > 1 \) for all \( l \in [n] \setminus \{i\} \). Subtracting \( \sum_{v \in V} \lambda_v = 1 \) and appropriate multiples of constraints \(-\lambda_v \leq 0\) from this constraint we obtain a scaling of (28), which shows the result. \( \square \)
A rational MIP formulation can be obtained by further requiring \( a^{i,j} \) and \( r_j \) to be rational. For instance, we can take integer points in \( S^2_+(r_j) \) for a small \( r_j \in \mathbb{Z} \) such that \( S^2_+(r_j) \) contains at least \(|I_j| + 1\) integer points. In particular, we may take \( r_j = 5^{2n} \) for \( \alpha := |(|I_j| - 1)/2 | \) to ensure \( S^2_+(r_j) \) contains enough integer points (e.g. [25, Theorem 278]). Hence, with this approach we may construct a version of (27) that has coefficients with sizes that are polynomial in \( \max_j |I_j| \) and hence in \( n \). This leads to the somewhat surprising fact that SOS2 constraints and constraints with disjoint \( T_i \)'s have a sharp (and encoding sharp) polynomial sized formulation with a constant (independent of the size of the constraint) number of embedding inequalities.

**Corollary 2.** If \( T \) is such that \( T_i \cap T_j = \emptyset \) for all \( i \neq j \), then
- \( \text{rc}_M (T) = 2 \), and
- \( 2 \leq \text{rc} (T) \leq 2 + |V| + n \).

Furthermore, the difference between the bounds on \( \text{rc} (T) \) is only due to variable bounds, and the upper bound is achieved by a formulation with only two equations.

**Corollary 3.** If \( T \) is an SOS2 constraint on \( V = [n+1] \), then
- \( 2 \leq \text{rc}_M (T) \leq 4 \), and
- \( 2 \leq \text{rc} (T) \leq 5 + 2n \).

Furthermore, the difference between the non-constant terms in the bounds on \( \text{rc} (T) \) is only due to variable bounds, and the upper bound is achieved by a formulation with only two equations.

Unfortunately, the following examples also show that the formulations fail to be ideal and their LP relaxation can have a large number of fractional extreme points. Hence, it is unlikely that it will provide a consistent computational advantage over existing small embedding formulations (e.g. (10) for the disjoint case and (35) for SOS2 constraints).

**Example 4.** Consider a disjoint constraint on \( V = [10] \) with \( T_i = \{ 2i - 1, 2i \} \) for \( i \in [5] \). The version of formulation (27) based on integral points in \( S^2_+(r_j) \) for \( r_j = 5^4 \) is given by

\[
25 (\lambda_1 + \lambda_2) + 18 (\lambda_3 + \lambda_4) + 10 (\lambda_5 + \lambda_6) + 5 (\lambda_7 + \lambda_8) + 1 (\lambda_9 + \lambda_{10}) \leq 25 y_1 + 18 y_2 + 10 y_3 + 5 y_5 + 1 y_5 \quad (30a)
\]

\[
0 (\lambda_1 + \lambda_2) + 1 (\lambda_3 + \lambda_4) + 5 (\lambda_5 + \lambda_6) + 10 (\lambda_7 + \lambda_8) + 18 (\lambda_9 + \lambda_{10}) \leq 0 y_1 + 1 y_2 + 5 y_3 + 10 y_4 + 18 y_5 \quad (30b)
\]

\[
\sum_{i=1}^{10} \lambda_i = 1 \quad (30c)
\]

\[
\sum_{i=1}^{5} y_i = 1 \quad (30d)
\]

\[
\lambda \in \mathbb{R}_{+}^{10} \quad (30e)
\]

\[
y \in \{0, 1\}^5 \quad (30f)
\]

The LP relaxation of formulation (30) has 80 fractional extreme points.

**Example 5.** Consider an SOS2 constraint on \( V = [9] \) and let \( I_1 = \{ 1, 3, 5, 7 \} \) and \( I_2 = \{ 2, 4, 6, 7 \} \). The version of formulation (27) based on integral points in \( S^2_+(r_j) \) for \( r_j = 5^4 \) is given by

\[
25 (\lambda_1 + \lambda_2) + 18 (\lambda_3 + \lambda_4) + 10 (\lambda_5 + \lambda_6) + 5 (\lambda_7 + \lambda_8) + 19 (\lambda_9) \leq 25 y_1 + 18 y_3 + 10 y_5 + 5 y_7 + 25 (y_2 + y_4 + y_6 + y_8) \quad (31a)
\]

\[
0 (\lambda_1 + \lambda_2) + 1 (\lambda_3 + \lambda_4) + 5 (\lambda_5 + \lambda_6) + 10 (\lambda_7 + \lambda_8) + 18 (\lambda_9) \leq 0 y_1 + 1 y_3 + 5 y_5 + 10 y_7 + 18 (y_2 + y_4 + y_6 + y_8) \quad (31b)
\]

\[
25 (\lambda_2 + \lambda_3) + 18 (\lambda_4 + \lambda_5) + 10 (\lambda_6 + \lambda_7) + 5 (\lambda_8 + \lambda_9) + 19 (\lambda_{10}) \leq 25 y_2 + 18 y_4 + 10 y_6 + 5 y_8 + 25 (y_1 + y_3 + y_5 + y_7) \quad (31c)
\]
Furthermore, (33) is faced defining if and only if

\[ \text{dim}(Q) \]

The LP relaxation of formulation (31) has 875 fractional extreme points.

5. **Bounds on embedding complexity for SOS2 constraints** In this section we conduct a detailed analysis of all embedding formulations for SOS2 constraints on \( V = [n+1] \). We begin with the following proposition that precisely characterizes such formulations and relates them to a family of hyperplanes associated to the encoding used to construct the formulation.

**Proposition 11.** For \( H : = \{ h^i \}_{i=1}^n \in \mathcal{H}_k(n) \) let \( c^i = h^{i+1} - h^i \) for \( i \in [n-1] \) and for \( b \in L(H) \setminus \{0\} \) let \( M(b) : = \{ y \in L(H) : b \cdot y = 0 \} \) be the linear (or central) hyperplane defined by \( b \) in \( L(H) \).

If \( \{ b^i \}_{i=1}^L \subseteq L(H) \setminus \{0\} \) is such that \( \{ M(b) \}_{i=1}^L \) is the set of linear hyperplanes spanned by \( \{ c^i \}_{i=1}^{n-1} \) in \( L(H) \), then \( Q(T,H) \) is equal to

\[
(b^i \cdot h^i) \lambda_i + \sum_{v=2}^{n} \min \{ b^i \cdot h^n, b^i \cdot h^{v-1} \} \lambda_v + (b^i \cdot h^n) \lambda_{n+1} \leq b^i \cdot y \quad \forall i \in [L]
\]

\[
-(b^i \cdot h^i) \lambda_i - \sum_{v=2}^{n} \max \{ b^i \cdot h^n, b^i \cdot h^{v-1} \} \lambda_v - (b^i \cdot h^n) \lambda_{n+1} \geq -b^i \cdot y \quad \forall i \in [L]
\]

\[
\sum_{v=1}^{n+1} \lambda_v = 1
\]

\[
\lambda_v \geq 0 \quad \forall v \in [n+1]
\]

\[
y \in L(H).
\]

Furthermore, \( \text{size}_M(Q(T,H)) = 2L, 2 \leq \text{size}_B(Q(T,H)) \leq n+1 \) and \( \text{size}_E(Q(T,H)) = 1 + k - \dim(H) \).

**Proof.** As usual (e.g. see proof of Lemma 6) and by Proposition 7, we may assume without loss of generality that any facet defining embedding inequality of \( Q(T,H) \) is of the form

\[ a \cdot \lambda \leq b \cdot y \quad (33) \]

with \( a \neq 0 \) and \( b \in L_1(H) : = \{ b \in L(H) : ||b|| = 1 \} \) where \( ||b|| \) is the euclidean norm. For each \( v \in [n+1], a \in \mathbb{R}^{n+1} \) and \( b \in \mathbb{R}^n \) let

\[
E_v(a,b) := \begin{cases} \{ i \in \{1\} : a_v = b \cdot h^i \} & v = 1 \\ \{ i \in \{v-1,v\} : a_v = b \cdot h^i \} & v \in [2,n] \\ \{ i \in \{n\} : a_v = b \cdot h^i \} & v = n+1 \end{cases}
\]

The set of extreme points of \( Q(T,H) \) supporting (33) (i.e. those that satisfy it at equality) is precisely

\[ X(a,b) := \{ (e^*,h^i) : v \in V, \quad i \in E_v(a,b) \} \].

Furthermore, (33) is faced defining if and only if \( X(a,b) \) is maximal with respect to inclusion (among all \( a \in \mathbb{R}^{n+1} \) and \( b \in L_1(H) \) for which (33) is an embedding inequality; c.f. Lemma 5). Hence, (33) is facet defining if and only if
1. \(|E_v(a,b)| \geq 1\) for all \(v \in \llbracket n+1 \rrbracket\), and
2. \(\{v \in \llbracket n+1 \rrbracket : |E_v(a,b)| = 2\}\) is maximal with respect to inclusion.

We have that \(|E_v(a,b)| = 2\) can only hold if \(v \in \llbracket 2, n \rrbracket\) and \(c^{v-1} \cdot b = 0\). Hence, condition 2 can only hold if \(I(b) := \{i \in \llbracket n-1 \rrbracket : c^i \cdot b = 0\}\) is maximal. Furthermore, if \(I(b)\) is maximal then the unique \(a \in \mathbb{R}^{n+1}\) such that \(a\) and \(b\) satisfies both conditions and for which \((33)\) is an embedding inequality is given by

\[
\begin{align*}
    a_1 &= b \cdot h^1, \\
    a_v &= \min \{b \cdot h^{v-1}, b \cdot h^v\} & v \in \llbracket 2, n \rrbracket, \\
    a_{n+1} &= b \cdot h^n.
\end{align*}
\]

We claim that there is a one-to-one correspondence between maximal sets \(I(b)\) for \(b \in L_1(H)\) and the hyperplanes spanned by \(\{c^i\}_{i=1}^{n-1}\) in \(L(H)\). Indeed, let \(M(b) := \{y \in L(H) : b \cdot y = 0\}\) be a hyperplane spanned by \(\{c^i\}_{i \in I(b)} \) for some \(I \subseteq \llbracket n-1 \rrbracket\). Without loss of generality we may assume that \(b \in L_1(H)\) and \(I = I(b)\). Now, because \(M(b)\) is a hyperplane in \(L(H)\) spanned by \(\{c^i\}_{i \in I(b)}\) we have that \(\dim \left( \{c^i\}_{i \in I(b)} \right) = \dim (H) - 1\). Then, because vectors in \(L_1(H)\) are non-zero, \(c^i \notin \text{span} \left( \{c^j\}_{j \in \llbracket n-1 \rrbracket \setminus I(b)} \right)\) for \(j \in \llbracket n-1 \rrbracket \setminus I(b)\) and \(\{c^i\}_{i=1}^{n-1} \subseteq L(H)\) we have that \(I(b)\) is maximal. For the converse let \(b \in L_1(H)\) be such that \(I(b)\) is maximal. Then, because \(\{c^i\}_{i=1}^{n-1}\) spans \(L(H)\) we have that \(\dim \left( \{c^i\}_{i \in I(b)} \right) = \dim (H) - 1\) and hence \(M(b)\) is a hyperplane spanned by \(\{c^i\}_{i \in I(b)}\).

To obtain the desired result it suffices to note that \(I(b) = I(-b)\) and to show that if \(I(b)\) is maximal, then \(b\) and \(-b\) yield different facets. The latter holds if there exists \(v \in \llbracket 2, n \rrbracket\) such that \(b \cdot h^{v-1} \neq b \cdot h^v\), which must hold for every \(b \in L_1(H)\). \(\square\)

Because Proposition 11 characterizes every possible embedding formulation for SOS2 we can use it to obtain the following lower bound on \(mc_M(T)\).

**Corollary 4.** If \(T\) is an SOS2 constraint on \(V = \llbracket n+1 \rrbracket\), then \(mc_M(T) \geq 2 \lceil \log_2 n \rceil\).

**Proof.** The number of (linear) hyperplanes spanned by \(\{c^i\}_{i=1}^{n-1}\) in \(L(H)\) is equal to the number of 1-flats of the (central) hyperplane arrangement \(\{\{b \in L(H) : c^i \cdot b = 0\}_{i=1}^{n-1}\}\) in \(L(H)\). Because \(\text{span} \left( \{c^i\}_{i=1}^{n-1} \right) = L(H)\) the number of such 1-flats is at least \(\dim (L(H)) = \dim (H)\). Because all elements of \(H\) are pairwise distinct we have that \(\dim (H) \geq \lceil \log_2 n \rceil\), which proves the result. \(\square\)

Proposition 11 also allows us to recover known formulations for SOS2 constraints. For instance, the following corollary characterizes the unary embedding formulation for SOS2 constraints considered in \([33, 38]\).

**Corollary 5.** If \(T\) is an SOS2 constraint on \(V = \llbracket n+1 \rrbracket\) and \(H\) is the unary encoding, then \(Q(H)\) is equal to

\[
\begin{align*}
    \sum_{i=l+2}^{n+1} \lambda_i &\leq \sum_{i=l+1}^{n} y_i & \forall l \in \llbracket n-1 \rrbracket \\
    \sum_{i=1}^{n} \lambda_i &\leq \sum_{i=1}^{n-1} y_i & \forall l \in \llbracket n-1 \rrbracket \\
    \sum_{i=1}^{n} \lambda_v &= 1 \\
    \lambda_1 &\geq 0 \\
    \lambda_{n+1} &\geq 0.
\end{align*}
\]

Furthermore, \(\text{size}_M (Q(H)) = 2(n-1)\), \(\text{size}_B (Q(H)) = 2\) and \(\text{size}_E (Q(H)) = 2\).
In addition, Proposition 11 can be used to recover and generalize a binary embedding-like formulation for SOS2 constraints introduced in [52], which we denote the logarithmic formulation for SOS2. To describe such formulation we need the following definition, which describes a special class of binary encodings where adjacent elements (in the order induced by the SOS2 constraints) only differ in one bit or coordinate.

**Definition 15.** For $h \in \mathbb{R}^k$ let $\text{supp}(h) := \{i \in [k] : h_i \neq 0\}$ its support and for $I, J \subseteq [k]$ let $I \oplus J := (I \setminus J) \cup (J \setminus I)$ be their symmetric difference. We say $H = \{h^i\}^n_{i=1} \in \mathcal{H}_{|\log_2 n|}(n)$ is a gray code if and only if

$$|\text{supp}(h^i) \oplus \text{supp}(h^{i+1})| = 1 \quad \forall i \in [n-1].$$

Gray codes exist for any $n$ including the case where $n$ is not a power of two [55]. This makes the logarithmic formulation valid for any value of $n$. However, as we will discuss in Section 6, the logarithmic formulation is technically an embedding formulation only when $n$ is a power of two. Fortunately, the following corollary of Proposition 11 adapts the logarithmic formulation to an embedding formulation for all $n$ (see [36] for an alternate derivation).

**Corollary 6.** If $\mathcal{T}$ is is an SOS2 constraint on $V = [n+1]$ and $H$ is a gray code, then $Q(H)$ is equal to

\[
\begin{align*}
  h^i_1 \lambda_1 + \sum_{v=2}^{n} \min \{ h^v_i, h^{v-1}_i \} \lambda_v + h^n_i \lambda_{n+1} &\leq y_i \quad \forall i \in \left\lceil \log_2 n \right\rceil \quad (35a) \\
  -h^i_1 \lambda_1 - \sum_{v=2}^{n} \max \{ h^v_i, h^{v-1}_i \} \lambda_v - h^n_i \lambda_{n+1} &\leq -y_i \quad \forall i \in \left\lceil \log_2 n \right\rceil \\
  \sum_{i=1}^{n+1} \lambda_v &= 1 \\
  \lambda_v &\geq 0 \quad \forall v \in [n+1]. \quad (35d)
\end{align*}
\]

Furthermore, $\text{size}_M(Q(H)) = 2 \lceil \log_2 n \rceil$, $2 \leq \text{size}_B(Q(H)) \leq n + 1$ and $\text{size}_E(Q(H)) = 1$.

When combined with the lower bound from Corollary 4 we see that the adapted logarithmic formulation is optimal with respect to embedding inequalities and essentially optimal (up to logarithmic terms) with respect to all inequalities.

**Corollary 7.** If $\mathcal{T}$ is is an SOS2 constraint on $V = [n+1]$, then

- $m_{c_M}(\mathcal{T}) = 2 \lceil \log_2 n \rceil$, and
- $n + 1 \leq xc(\mathcal{T}) \leq xmc(\mathcal{T}) \leq mc(\mathcal{T}) \leq n + 1 + 2 \lceil \log_2 n \rceil$.

Furthermore, the upper bounds are achieved by a formulation with only one equation.

**Proof.** The bounds on the embedding complexity are direct from Lemma 2 and Corollaries 4 and 6. The lower bound on the extension complexity is direct by noting that the slack matrix of $\Delta^V$ is a $|V| \times |V|$ identity plus two zero rows, which has a non-negative rank of $|V|$. □

Finally, note that the smallest alternative ideal formulation for SOS2 constraints is the extended formulation from Proposition 4, which has size $2n$. Hence, Corollary 7 also improves on the best known upper bound on the extended embedding complexity of SOS2 constraints. Furthermore, it shows that a careful selection of the encoding can lead to embedding formulations that are smaller than the best known extended formulation. In the following subsection we study the sensitivity of the size of the embedding formulation to this encoding selection.
5.1. Size distribution for binary encodings. If \( n = 2^k \) for some \( k \in \mathbb{Z} \) and \( H \in \mathcal{H}_k(n) \), then Proposition 11 shows that the number of embedding inequalities of \( Q(H) \) is upper bounded by \( 2^{\binom{n-1}{k-1}} \). The following result suggests that this upper bound may be nearly achieved.

**Definition 16.** Let \( n = 2^k \) for some \( k \in \mathbb{Z} \). We say \( H = \{ h^i \}_{i=1}^n \subset \mathcal{H}_k(n) \) is an **anti-gray code**\(^{††} \) if and only if

\[
\begin{align*}
\supp(h^{2i-1}) \oplus \supp(h^{2i}) &= n & \forall i \in [n/2] \\
\supp(h^{2i}) \oplus \supp(h^{2i+1}) &= n - 1 & \forall i \in [n/2 - 1].
\end{align*}
\]

Anti-gray codes exist for all \( k \) and can easily be constructed from gray codes (e.g. [40]). The following lemma refines Proposition 11 for embedding formulations constructed from anti-gray codes.

**Lemma 8.** Let \( n = 2^k \) for some \( k \in \mathbb{Z} \) and \( H \subset \mathcal{H}_k(n) \) be an anti-gray code. Then the number of embedding inequalities of \( Q(H) \) is equal to twice the number of affine hyperplanes spanned by \( \{0, 1\}^{k-1} \).

**Proof.** Let \( c^i = h^{i+1} - h^i \) for \( i \in [n-1] \). Because \( H \) is an anti-gray code there exist \( I \subset [n-1] \) with \( |I| = 2^{k-1} \) such that \( c^i \in \{-1, 1\}^k \) for all \( i \in I \). In addition, because \( h^i \neq h^j \) for \( i \neq j \) we have that \( c^i \neq -c^i \) for all \( i, j \in I \). Hence for all \( s \in \{-1, 1\}^k \) there exist \( i \in I \) such that \( s = c^i \) or \( s = -c^i \). The result then follows from Proposition 11 by noting that \( \{c^i\}_{i \in I} \) and \( \{\pm c^i\}_{i \in I} \) span the same set of linear hyperplanes and that the number of linear hyperplanes spanned by \( \{-1, 1\}^k \) is equal to the number of affine hyperplanes spanned by \( \{0, 1\}^{k-1} \) (e.g. [15]) \( \square \)

It is believed that the number of affine hyperplanes spanned by \( \{0, 1\}^{k-1} \) is close to its trivial upper bound of \( \binom{n/2}{k-1} \) for \( n = 2^k \) (e.g. [1]). Both this upper bound and the embedding inequality bound of \( 2\binom{n-1}{k-1} \) grow roughly as \( n^{\log_2 n} \), which suggests that the worst case for the number of embedding inequalities of a binary encoded formulation is quasi-polynomial in \( n \). Hence, it seems like an unfortunate selection of the specific binary encoding can lead to a formulation that is significantly larger than the upper bound from Corollary 4 or even the size of the unary encoded formulation from Corollary 5. Because its link with the number of hyperplanes spanned by \( \{0, 1\}^{k-1} \) (or \( \{-1, 0, 1\}^{k-1} \), understanding the typical size of a binary encoded embedding formulation for SOS2 constraints may prove extremely challenging (e.g. [53]). For this reason we only pursue a simple empirical study of the distribution of sizes for these formulations. For this study we selected \( k \in [3, 15] \) and calculated the number of embedding inequalities for randomly selected binary encodings (the ones associated to a random permutation of \( \{0, 1\}^k \)). For \( n = 3 \) we considered all 40320 possible encodings, while for \( k = 4, 5 \) we only used a random sample of 10000 encodings and for \( k = 6 \) we only used a random sample of 1000 encodings (calculating the formulation sizes for \( k = 6 \) was already computational intensive). The results of this study are presented in Figure 1. The figure presents histograms for the number of embedding inequalities for each \( k \), together with the trivial upper bound of \( 2\binom{n-1}{k-1} \) (depicted by the solid red line), the number of embedding inequalities of the unary encoded embedding formulation (depicted by the dotted blue line) and the number of embedding inequalities of the optimal binary encoded formulation (depicted by the dashed green line). The figure shows that the typical number of embedding inequalities of a binary encoded formulation seems to be much closer to the upper bound and suggests that a randomly selected encoding may often lead to a formulation that is significantly larger than even the unary encoded formulation. Hence a careful encoding selection appears crucial to obtain a small formulation.

\( \dagger \) The class of codes obtained by switching \( n \) and \( n - 1 \) in this definition is sometimes also referred to as anti-gray code.

\( \dagger \dagger \) We have \( \Omega \left( n^{(1 - \varepsilon) \log_2 n} \right) = \binom{n/2}{k-1} \leq \binom{n-1}{k-1} \leq n^{\log_2 n} \) for all \( \varepsilon > 0 \).
6. Redundant Formulations  
Section 5.1 suggests that the size of an embedding formulation can be extremely sensitive to the encoding selection. For SOS2 constraints, the detailed analysis of Proposition 11 allows for a direct construction of the smallest embedding formulation. However, this analysis is strongly dependent on the simple structure of SOS2 constraints. Fortunately, for $n$ a power of two, a procedure introduced in [52] can be used to construct the smallest embedding formulation for SOS2 constraints without the detailed analysis of Proposition 11. This approach was denoted the independent branching scheme and can be used to construct small ideal formulations for other families of polytopes that fit Definition 9 (i.e. faces of a simplex). The resulting formulations were shown to provide a significant computational advantage in [52] and [51]. However, the resulting formulations do not always fit Definitions 2 or 5 (i.e. our formal definitions of formulation and embedding formulation). For instance, consider the following example from [36].

**Example 6.** Let $T$ be a SOS2 constraint on $V = [4]$. If $H := \{h^1\}_{i=1}^3$ for $h^1 = (0, 0)^T$, $h^2 = (1, 0)^T$ and $h^3 = (1, 1)^T$, then $Q(T, H)$ is equal to

\[
\begin{align*}
\lambda_3 + \lambda_4 & \leq y_1 \\
-\lambda_2 - \lambda_3 - \lambda_4 & \leq -y_1 \\
\lambda_4 & \leq y_2 \\
-\lambda_3 - \lambda_4 & \leq -y_2 \\
\sum_{i=1}^{4} \lambda_i &= 1 \\
\lambda_i & \geq 0 \quad \forall i \in [4].
\end{align*}
\]

(36a) (36b) (36c) (36d) (36e) (36f)

In contrast, the LP relaxation of the independent branching formulation from [52] is given by

\[
\begin{align*}
\lambda_3 & \leq y_1 \\
-\lambda_2 - \lambda_3 - \lambda_4 & \leq -y_1 \\
\lambda_4 & \leq y_2 \\
-\lambda_3 - \lambda_4 & \leq -y_2 \\
\end{align*}
\]

(37a) (37b) (37c) (37d)
\[
\sum_{i=1}^{4} \lambda_i = 1 \quad \text{(37e)}
\]
\[
\lambda_i \geq 0 \quad \forall i \in [4]. \quad \text{(37f)}
\]

Both (36) and (37) yield ideal formulations for SOS2 constraint on \( V = [4] \), but (36) is strictly contained in (37). Indeed we can check that \((\lambda, y) = (0, 0, 0, 1, 0, 1)^T\) is feasible for (37), but not for (36).

Example 6 may seem contradictory at first: we have two ideal formulations for the same set, but one is strictly contained in the other. However, this apparent contradiction is resolved by noting that (36) and (37) do not yield formulations for the same set in the sense of Definition 2. More specifically, (36) yields a valid formulation for SOS2 constraints \( V = [4] \) in the sense of Definition 2. However, (37) only yields a valid formulation in a broader sense, as it introduces some redundancy that makes it fail the requirements of Definition 2. This redundancy has to do with the values of \( y \) for which \((\lambda, y)\) is feasible for (37) when \( \lambda = (0, 0, 0, 1)^T \). We certainly obtain a feasible \((\lambda, y)\) when \( y = (0, 0)^T = h^1 \), which fits Definition 2. However, we also obtain a feasible \((\lambda, y)\) when \( y = (0, 1)^T \notin \{h^1, h^2, h^3\} \), which does not fit Definition 2. Refining this analysis we can check that (37) is actually equal to \( Q(\mathcal{T}, \mathcal{H}) \) for \( \mathcal{T} = \{\mathcal{T}_i\}_{i=1}^{n} \) and \( \mathcal{H} = \{\mathcal{H}_i\}_{i=1}^{4} \) with \( \mathcal{T}_1 = \{1, 2\}, \mathcal{T}_2 = \{2, 3\}, \mathcal{T}_3 = \{3, 4\}, \mathcal{T}_4 = \{4\}, \mathcal{H}_1 = (0, 0)^T, \mathcal{H}_2 = (1, 0)^T, \mathcal{H}_3 = (1, 1)^T \) and \( \mathcal{H}_4 = (0, 1)^T \). The reason (37) yields a valid formulation for SOS2 constraints on \( V = [4] \) stems from \( \mathcal{T}_i = T_i \) for \( i \in [3] \) and \( \mathcal{T}_4 \subseteq \mathcal{T}_3^{36} \). This implies \( \bigcup_{i=1}^{4} P(T_i) = \bigcup_{i=1}^{4} P(\mathcal{T}_i) \) and hence (37) is a valid formulation for SOS2 constraints because it is a valid formulation for an alternative family of polyhedra which has the same union as those associated to the SOS2 constraints. The following definition formalizes this property and extends Definitions 2 and 5 to allow for such valid redundancy. This definition also allows for other relations between the original and alternative family of polyhedra (e.g. some elements of the original family are unions of the alternative family) and focuses only on the ideal version of the formulations (i.e. it actually extends Definition 5). The analog extension of Definition 2 is straightforward.

**Definition 17 (Redundant Embedding Formulation).** Let \( \mathcal{P} := \{P^n_i\}_{i=1}^{n} \) be a finite family of non-empty polyhedra in \( \mathbb{R}^d \) and \( H := \{h^m_i\}_{i=1}^{m} \in \mathcal{H}(m) \) for \( m > n \). We say a polyhedron \( R \) is a redundant embedding formulation for \( \bigcup_{i=1}^{n} P^i \) if there exist a family of non-empty polyhedra \( \mathcal{P} = \{P^i\}_{i=1}^{m} \) such that

1. \( \bigcup_{i=1}^{m} P^i = \bigcup_{i=1}^{n} P^i \); and
2. \( R = Q(\mathcal{P}, H) \).

We let \( \text{rmc}(\mathcal{P}) \) and \( \text{rmc}_M(\mathcal{P}) \) be the analogs of \( \text{mc}(\mathcal{P}) \) and \( \text{mc}_M(\mathcal{P}) \) for redundant embedding formulations.

The following proposition shows that the independent branching approach from [52] yields a redundant embedding formulation.

**Proposition 12.** Let \( \{B_{k,0}, B_{k,1}\}_{k=1}^{L} \) with \( B_{k,0}, B_{k,1} \subseteq V \) be such that

1. \( B_{k,0} \cap B_{k,1} = \emptyset \) for all \( k \in [L] \),
2. For all \( i \in [n] \) there exist \( s \in \{0, 1\}^L \) such that \( \bigcap_{k=1}^{L} B_{k,s_k} = T_i \), and
3. For all \( s \in \{0, 1\}^L \) there exist \( i \in [n] \) such that \( \bigcap_{k=1}^{L} B_{k,s_k} \subseteq T_i \).

Then \( R \) defined by

\[
\sum_{v \in V \setminus B_{k,0}} \lambda_v \leq 1 - y_k \quad \forall k \in [L] \quad \text{(38a)}
\]

\(36\)The similarity with (37) follows from this and \( \mathcal{H}_i = h^i \) for \( i \in [3] \).
\[ \sum_{v \in V \setminus B_{k,1}} \lambda_v \leq y_k \quad \forall k \in [L] \quad (38b) \]
\[ \sum_{v \in V} \lambda_v = 1 \quad (38c) \]
\[ \lambda_v \geq 0 \quad \forall v \in V \quad (38d) \]
\[ y \in \mathbb{R}^L \quad (38e) \]

is a redundant embedding formulation for \( \bigcup_{i=1}^n P(T_i) \).

**Proof.** Let \( \mathcal{T} = \{ \mathcal{T}_s \}_{s \in \{0,1\}^L} \) with \( \mathcal{T}_s = \bigcap_{k=1}^L B_{k,s_k} \) and \( H = \{ s \}_{s \in \{0,1\}^L} \). Following the proof of Theorem 1 in [51] we have that \((R, H)\) is an ideal formulation of \( \bigcup_{i=1}^m \mathcal{P}^i \). Then by Proposition 1 \( R = Q(\mathcal{T}, H) \). \( \Box \)

Formulation (38) provides a way to construct binary encoded (redundant) embedding formulations that is much simpler than analyzing \( Q(H) \) for all encodings \( H \). This approach was used in [52] to construct formulations for a family of polyhedra associated with piecewise linear functions of two variables. However, the resulting formulations were restricted to a special class of piecewise linear functions. In the following section we show how the procedure can be adapted to a broader class of functions and study how the redundancy in Definition 17 can help reduce formulation sizes.

### 6.1. Triangulations and piecewise linear functions

MIP formulations for multivariate piecewise linear functions can be constructed from disjunctions that fit Definition 9 (e.g. [33, 52, 51]). For simplicity, we consider formulations for piecewise linear functions of two variables defined on grid triangulations on \( [m+1]^2 \) such as those depicted in Figure 2 (for more general cases see [33, 51]). If \( f : [1, m+1]^2 \to \mathbb{R} \) is continuous in \( [1, m+1]^2 \) and affine in each triangle of the triangulation (e.g. for the triangulation in Figure 2(a) it is affine in \( \text{conv} \{\{(1,1),(1,2),(2,1)\}\} \)), then we can construct a formulation of the graph or epigraph of \( f \) through a specific disjunction associated to the triangulation. To define this disjunction note that for a fixed \( m \) a triangulation is uniquely determined by how each square in the grid associated to \( [m+1]^2 \) is divided into two triangles. The following definition formalizes this and simultaneously introduces the disjunction used to construct the formulation.

**Definition 18 (Grid Triangulation).** \( V = [m+1]^2, \mathcal{T} = \{ S_r^{gr}, T_r^{gr} \}_{r \in [m]^2} \) for \( \{ g_r \}_{r \in [m]^2} \subseteq \{1, 2\} \) such that

![Figure 2. Two grid triangulations of V = [3]^2.](image)
achieve this we need the following definition. To use this, we can construct a redundant embedding formulation (which happens to also be the embedding formulation) for a triangulation known as the union-jack, which corresponds to the embedding formulation) for a triangulation known as the union-jack, which corresponds to the union-jack triangulation is not without consequence as other triangulations are sometimes preferable (e.g. [47]). For this reason we now show how the independent branching approach can be used to construct a redundant embedding formulation for a wider class of triangulations. To achieve this we need the following definition.

**Definition 19 (Coloring of a Triangulation).** Let

\[
V_1 = \left\{ (i, j) \in [m + 1]^2 : i \text{ and } j \text{ have equal parity} \right\}
\]

and

\[
V_2 = \left\{ (i, j) \in [m + 1]^2 : i \text{ and } j \text{ have different parity} \right\}.
\]

We say a pair of labellings \( C_l : V_l \rightarrow \{0, 1\} \) for \( l \in \{1, 2\} \) is a coloring of the grid triangulation associated to \( V = [n + 1]^2 \) and \( \{g_r\}_{r \in [m]^2} \subseteq \{1, 2\} \) if and only if for all \( l \in \{1, 2\} \)

1. \( C_l(i, j) \neq C_l(i + 1, j + 1) \) for all \( (i, j) \in V_l \cap [n - 1]^2 \) such that \( g(i, j) = 2 \)
2. \( C_l(i + 1, j) = C_l(i, j + 1) \) for all \( (i, j) \in V_l \cap [n - 1]^2 \) such that \( g(i, j) = 2 \),
3. \( C_l(i + 1, j) \neq C_l(i, j + 1) \) for all \( (i, j) \in V_l \cap [n - 1]^2 \) such that \( g(i, j) = 1 \), and
4. \( C_l(i, j) = C_l(i + 1, j + 1) \) for all \( (i, j) \in V_l \cap [n - 1]^2 \) such that \( g(i, j) = 1 \).

Figure 3(a) shows a coloring for the triangulation in Figure 2(a). Elements in \( V_1 \) are marked with circles, elements in \( V_2 \) are marked with diamonds and the associated disjunctions. Using a coloring of a triangulation we obtain the following generalization of the formulation from [52], which follows directly from Proposition 12.

**Proposition 13.** Let

- \( C_l : V_l \rightarrow \{0, 1\} \) for \( l \in \{1, 2\} \) be a coloring of a grid triangulation, and
- \( H = \left\{ (h_i)_{i=1}^m \subseteq \{0, 1\}^{[\log_2 m]} \right\} \) be a gray code.

If \( |C_1(V_1)| = |C_2(V_2)| = 2 \), then a redundant embedding formulation for \( \bigcup_{i=1}^n P(T_i) \) is given by

\[
\sum_{v \in V_l : C_l(v) = 0} \lambda_v \leq 1 - y^l \quad \forall l \in \{1, 2\}, \tag{39a}
\]

\[
\sum_{v \in V_l : C_l(v) = 1} \lambda_v \leq y^l \quad \forall l \in \{1, 2\}, \tag{39b}
\]

\[
\sum_{i=1}^m \left( -h_k \lambda_{(1,i)} - \sum_{j=2}^m \max \left\{ h_j^m, h_k^{-1} \right\} \lambda_{(j,i)} - h_k^m \lambda_{(m+1,i)} \right) \leq -y_k^3 \quad \forall k \in [\lceil \log_2 n \rceil] \tag{39c}
\]

\[
\sum_{i=1}^m \left( h_k \lambda_{(1,i)} + \sum_{j=2}^m \min \left\{ h_j^m, h_k^{-1} \right\} \lambda_{(j,i)} + h_k^m \lambda_{(m+1,i)} \right) \leq y_k^3 \quad \forall k \in [\lceil \log_2 m \rceil] \tag{39d}
\]
Finally, all upper bounds are achieved by 

\[ (39) \]

If the triangulation has a coloring with 

\[ \left| C_1(V_1) \right| + \left| C_2(V_2) \right| = 3, \] then an embedding formulation for \( \bigcup_{i=1}^{n} P(T_i) \) is given by (39) with \( y^l \) and constraints (39a)–(39b) for \( l \) such that \( \left| C_i(V_i) \right| = 1 \) removed.

**Proof.** From Proposition 12 by noting that (39c) and (39e) are equivalent to

\[
\begin{align*}
\sum_{i=1}^{m} \left( -h_k^1 \lambda_{i,1} \right) + \sum_{j=2}^{m} \left( h_k^j \lambda_{i,j} \right) & \leq -y_k^4 \quad \forall k \in \left[ \left\lfloor \log_2 n \right\rfloor \right] (39e) \\
\sum_{i=1}^{m} \left( h_k^j \lambda_{i,j} \right) + \sum_{j=2}^{m} \left( h_k^j \lambda_{i,j} \right) & \leq y_k^4 \quad \forall k \in \left[ \left\lfloor \log_2 m \right\rfloor \right] (39f)
\end{align*}
\]

where \( \left\{ y^1 \right\}^{m}_{i=1} \subseteq \{0, 1\}^{\left\lfloor \log_2 m \right\rfloor} \) is such that \( g_k^i = 1 - h_k^i \) for all \( i \in [m], k \in \left[ \left\lfloor \log_2 m \right\rfloor \right] \). \( \Box \)

The following direct corollary shows that Proposition 13 yields formulations for piecewise linear functions with sizes that are optimal up to logarithmic terms.

**Corollary 8.** Let \( \mathcal{T} = \{T_i\}^{n}_{i=1} \) for \( n = m^2 \) be a grid triangulation on \( V = [m + 1]^2 \). If the triangulation has a coloring with \( \left| C_1(V_1) \right| = \left| C_2(V_2) \right| = 2, \) then

\[
\text{rmc}_M(\mathcal{T}) \leq 4 \left[ \log_2 m \right] + 4 = 4 \left[ \log_2 \sqrt{n/2} \right] + 4
\]

and

\[
(m + 1)^2 \leq \text{xc}(\mathcal{T}) \leq \text{rmc}(\mathcal{T}) \leq (m + 1)^2 + 4 \left[ \log_2 m \right] + 4
\]

or equivalently

\[
\left( \sqrt{n/2} + 1 \right)^2 \leq \text{xc}(\mathcal{T}) \leq \text{rmc}(\mathcal{T}) \leq \left( \sqrt{n/2} + 1 \right)^2 + 4 \left[ \log_2 \sqrt{n/2} \right] + 4.
\]

If the triangulation has a coloring with \( \left| C_1(V_1) \right| + \left| C_2(V_2) \right| = 3, \) then

\[
\text{mc}_M(\mathcal{T}) \leq 4 \left[ \log_2 m \right] + 2 = 4 \left[ \log_2 \sqrt{n/2} \right] + 2
\]

and

\[
(m + 1)^2 \leq \text{xc}(\mathcal{T}) \leq \text{mc}(\mathcal{T}) \leq (m + 1)^2 + 4 \left[ \log_2 m \right] + 2
\]

or equivalently

\[
\left( \sqrt{n/2} + 1 \right)^2 \leq \text{xc}(\mathcal{T}) \leq \text{mc}(\mathcal{T}) \leq \left( \sqrt{n/2} + 1 \right)^2 + 4 \left[ \log_2 \sqrt{n/2} \right] + 2.
\]

Finally, all upper bounds are achieved by (39), which only has one equation.
As a point of comparison consider the following direct corollary from results in [33], which gives the size of the unary encoded embedding formulation for grid triangulations.

**Corollary 9.** Let \( V = [m+1]^2 \), \( T \) a grid triangulation on \( V = [m+1]^2 \) and \( H \) be the unary encoding. Then

\[
\text{size}_M(Q(T, H)) = \left( \frac{2m}{m} \right) = \left( \frac{2\sqrt{n/2}}{\sqrt{n/2}} \right).
\]

and

\[
\text{size}(Q(T, H)) = \left( \frac{2m}{m} \right) + (m+1)^2 = \left( \frac{2\sqrt{n/2}}{\sqrt{n/2}} \right) + \left( \sqrt{n/2} + 1 \right)^2.
\]

**Proof.** Direct from Proposition 10 in [33] and the comments before its statement, and Proposition 1. \( \square \)

We can see that the independent formulation is significantly smaller that the unary encoding formulation. In particular, the number of embedding inequalities of the unary encoded formulation is superpolynomial in \( n \), while the independent branching formulation is only linear in \( n \). Furthermore, the smallest alternative ideal formulation for grid triangulations is the extended formulation from Proposition 4. This formulation has size \( 3n \), which is about six times the size of the non-extended embedding formulation from Corollary 8. Hence, Corollary 8 also improves on the best known upper bound on the extended embedding complexity for the grid triangulations it considers. Still, while Corollary 8 can deal with triangulations beyond the union-jack, it is easy to construct formulations for which it cannot yield a valid formulation.

We end this section with an example of a redundant embedding formulation from Corollary 8 that does not yield the embedding formulation. This example suggests that redundancy may be necessary to construct the smallest possible ideal formulation.

**Example 7.** For \( m = 2 \) consider the grid triangulation depicted in Figure 3 with \( g_r = 2 \) for all \( r \in [2]^2 \). One coloring of this triangulation is given by

\[
\begin{align*}
C_1(v) &: = 0 \quad \forall v \in \{(1,1), (3,3)\} \\
C_1(v) &: = 1 \quad \forall v \in \{(1,3), (2,2), (3,1)\} \\
C_2(v) &: = 0 \quad \forall v \in \{(1,2), (2,1)\} \\
C_2(v) &: = 1 \quad \forall v \in \{(2,3), (3,2)\},
\end{align*}
\]

which is illustrated in Figure 3(a) (Circles correspond to vertices in \( V_1 \) and diamonds correspond to vertices \( V_2 \)). Formulation (39) for this coloring is given by

\[
\begin{align*}
\lambda_{(1,1)} + \lambda_{(3,3)} &\leq 1 - y_1^4 \quad (40a) \\
\lambda_{(1,3)} + \lambda_{(2,2)} + \lambda_{(3,1)} &\leq y_1^2 \quad (40b) \\
\lambda_{(1,2)} + \lambda_{(2,1)} &\leq 1 - y_2^2 \quad (40c) \\
\lambda_{(2,3)} + \lambda_{(3,2)} &\leq y_2^2 \quad (40d) \\
\lambda_{(1,1)} + \lambda_{(1,2)} + \lambda_{(1,3)} &\leq 1 - y_3^4 \quad (40e) \\
\lambda_{(3,1)} + \lambda_{(3,2)} + \lambda_{(3,3)} &\leq y_3^2 \quad (40f) \\
\lambda_{(1,1)} + \lambda_{(2,1)} + \lambda_{(3,1)} &\leq 1 - y_4^3 \quad (40g) \\
\lambda_{(1,3)} + \lambda_{(2,3)} + \lambda_{(3,3)} &\leq y_4^2 \quad (40h) \\
\sum_{v \in V} \lambda_v &= 1 \quad (40i) \\
\lambda &\in \mathbb{R}_+^V \\
y_l &\in \{0, 1\} \quad \forall l \in \{1, 2\}, \quad (40j), (40k) \\
y^3, y^4 &\in \{0, 1\}. \quad (40l)
\end{align*}
\]
If we identify the coordinates of \( h_j^i \) for \( j \in \{1, \ldots, 4\} \) with \( y_1^i, y_2^i, y_3^i \) and \( y_4^i \) respectively, we can see that this redundant formulation has an assignment of the encoding to the elements in \( T \) given by

\[
\begin{align*}
  h^1 &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad T_1 = \{(1, 1), (1, 2), (2, 1)\}; & h^2 &= \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad T_2 = \{(1, 2), (2, 1), (3, 3)\}; \\
  h^3 &= \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad T_3 = \{(1, 2), (1, 3), (2, 2)\}; & h^4 &= \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad T_4 = \{(1, 3), (2, 3), (2, 2)\}; \\
  h^5 &= \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \quad T_5 = \{(2, 1), (2, 2), (3, 1)\}; & h^6 &= \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad T_6 = \{(2, 2), (3, 2), (3, 1)\}; \\
  h^7 &= \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad T_7 = \{(2, 2), (2, 3), (3, 2)\}; & h^8 &= \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad T_8 = \{(3, 2), (3, 2), (3, 3)\}.
\end{align*}
\]

This is illustrated in Figure 3(a) where the encoding is written inside the corresponding triangle. However, the formulation also augments \( T \) to \( \bar{T} \) by adding the single vertex elements given by

\[
\begin{align*}
  h^9 &= \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad T_9 = \{(1, 1)\}; & h^{10} &= \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad T_{10} = \{(2, 2)\}; & h^{11} &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad T_{11} = \{(1, 2)\}; \\
  h^{12} &= \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \quad \bar{T}_{12} = \{(3, 2)\}; & h^{13} &= \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \bar{T}_{13} = \{(2, 1)\}; & h^{14} &= \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \bar{T}_{14} = \{(3, 2)\}; \\
  h^{15} &= \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \bar{T}_{15} = \{(2, 2)\}; & h^{16} &= \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \quad \bar{T}_{16} = \{(3, 3)\}.
\end{align*}
\]

This is illustrated in Figure 3(b) where the encoding is written on top of the corresponding vertex.

The number of embedding inequalities of formulation (40) is 8, while

\[
\text{size}_M \left( Q \left( T, \{ h^i \}_{i=1}^8 \right) \right) = 19.
\]

Furthermore, every embedding formulation of \( T \) with \( H \subseteq \{0, 1\}^3 \) has at least 9 embedding inequalities. Hence, redundancy may be necessary to obtain the smallest possible (redundant or not) embedding formulation.

7. Further relations between complexity measures and future work We have shown that a careful selection of the encoding can yield embedding formulations that are smaller than alternative traditional formulations. Some of these formulations have been shown to provide a significant computational advantage \([33, 52, 51]\). Hence, an interesting direction of future work is to
understand how to construct such small formulations for a wider class of polyhedra. An alternative to directly constructing such small embedding formulations is to exploit redundancy through procedures such as the independent branching approach. We have shown that redundancy may be necessary to produce the smallest possible formulation. However, it is still not clear why redundancy seems to help reduce formulation sizes. Understanding this could be useful to construct redundant embedding formulations for a wider classes of polyhedra such as general grid triangulations. A final practical direction of research is understanding what is the price of requiring formulations to be ideal. We have shown that the number of embedding inequalities can be significantly reduced, but at the cost of large number of fractional extreme points and possible numerical stability. While unlikely, such formulations could still provide a computational advantage in some cases. Alternatively, it would be interesting to see if the trade-off between number of embedding inequalities and fractional extreme points can be balanced to obtain formulations that do provide a computational advantage.

From a theoretical perspective, there are still many open questions concerning the complexity measures for unions of polyhedra. In particular, we still do not have a complete picture of the relation between these measures. A summary of the relations that we do know about is depicted in Figure 4. Figure 4(a) shows the relationship between all non-redundant complexity measures and Figure 4(b) shows the relationship between the same complexity measures and the size of the Minkowski sum, which we denote by $\text{k}c(\mathcal{P}) := \text{size}(P_1 \ldots + P_n)$. For simplicity we exclude the redundant versions of the complexity measures as the only non-trivial relation we are aware of it the one suggested by Example 7. In Figure 4, a single filled arrow going from complexity $A$ to complexity $B$ indicates that $A \leq B$ for all families of polyhedra. A double filled arrow indicates that $A \leq B$ for all family of polyhedra and that there exists a family for which $A < B$. A double hollow arrow going from $A$ to $B$ indicates that there exists a family of polyhedra for which $A < B$, but we do not know if $A \leq B$ in general. In other words, a double hollow arrow going from $A$ to $B$ indicates that $A \geq B$ does not hold in general (i.e. the opposite single filled arrow cannot happen). In particular, if we have a double hollow arrow from $A$ to $B$ and from $B$ to $A$ the measures are incomparable. Labels C.P., A.P. and $\Delta$ indicate which specific family of polyhedra yields a strict inequality as we now detail.

The first class of polyhedra we consider yields strict inequalities for the hull complexity. This class corresponds to $P^i := \{x \in \mathbb{R}^n : -1 \leq x_i \leq 1, \ x_j = 0 \ \forall j \in [n] \setminus \{i\} \}$ for all $i \in [n]$, for which we have $\text{conv}(\bigcup_{i=1}^n P^i) = \{x \in \mathbb{R}^n : \|x\|_1 \leq 1\}$ (usually denoted the cross polytope) and hence $\text{k}c(\mathcal{P}) = 2^n$. An non-extended ideal formulation of size $3n$ for $\bigcup_{i=1}^n P^i$ is given in Example 3 of [50],
which shows the inequality with $xrc(\mathcal{P})$, $rc(\mathcal{P})$, $xmc(\mathcal{P})$ and $mc(\mathcal{P})$. Finally, $P^1 + \ldots + P^n = \{x \in \mathbb{R}^n : \|x\|_\infty \leq 1\}$ and hence $kc(\mathcal{P}) = 2n$, which shows the last strict inequality.

The second class of polyhedra we consider yields the strict inequality between the Minkowski and hull complexities and all strict inequalities for the embedding complexity. This class corresponds to the antiprisms studied in Example 2, for which we have already stated the corresponding strict inequalities in Section 2.2. Note that, except for the inequality between embedding and Minkowski complexities, all these strict inequalities are between a polynomial and an exponentially growing function. In contrast, the difference between the embedding and Minkowski complexities is only a constant.

The final class of polyhedra we consider yields the second strict inequality between the Minkowski and embedding complexities. This class corresponds to the grid triangulations for which there exist a coloring that satisfies $|C_1(V_1)| + |C_2(V_2)| = 3$. Using Corollary 9 and the Cayley trick we can show that

$$kc(\mathcal{P}) = \left(\frac{2\sqrt{n/2}}{\sqrt{n/2}}\right)$$

for this family. The strict inequality then follows from Corollary 8.

The challenge with most unknown relations stems from trade-offs between the requirements associated to the specific complexity measures. For instance, the trade-off for hull and relaxation complexities is between strength and actually being a formulation. Similarly, the trade-off for extended embedding and relaxation complexities is between strength and being a non-extended formulation. Fortunately, in some cases these trade-offs can be distilled to a specific geometric question. For instance, determining if there is a family of polyhedra for which the extension complexity is strictly smaller that the embedding complexity reduces to showing that the extension complexity of the family strictly increases for any valid embedding. Understanding the feasibility of this condition may be possible by adapting known extension complexity techniques. A final interesting question is to understand how the relaxation complexities change when we require the formulations to be sharp, but not ideal. Some interesting results in this direction can be found in [5, 8, 27].

Acknowledgments. This research was partially supported by the National Science Foundation under grant CMMI-1351619.
References


Appendix. Additional Examples

The following example shows that non-sharp formulations do not necessarily imply equations, bounds or encoding constraints.

Example 8. Consider the formulation of 3-dimensional SOS2 constraints given by

\[
\begin{align*}
y_1 \leq & \lambda_1 + \lambda_2 + \lambda_3 \leq 2 - y_1 \\
1 - y_1 \leq & \lambda_1 + \lambda_2 + \lambda_3 \leq 1 + y_1 \\
y_1 \leq & \lambda_1 + \lambda_2 \\
1 - y_1 \leq & \lambda_2 + \lambda_3 \\
y_1 - 1 \leq & \lambda_1 \\
y_1 - 1 \leq & \lambda_2 \\
-y_1 \leq & \lambda_2 \\
-y_1 \leq & \lambda_3 \\
-1 \leq & 2y_1 \\
2y_1 \leq & 3 \\
y_1 \in & \mathbb{Z}.
\end{align*}
\]

The LP relaxation of this formulation does not imply \(\sum_{i=1}^{3} \lambda_i = 1, \lambda_i \geq 0 \) or \(y_1 \in [0, 1]\).

The following example illustrates what can happen when the conditions of Lemma 6 and Proposition 7 do not hold.

Example 9. For \(V = \{1, 2\}, n = 2, T_1 = \{1, 2\}, T_2 = \{1\}, h^1 = 0, h^2 = 1\) we have that \(v = 1\) is not a cut vertex, but \(\lambda_1 \geq 0\) is not facet defining for \(Q(T, H)\). We also have that \(y_1 \geq 0\) is facet defining.