

Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and should not be used to distribute the papers in print or online or to submit the papers to another publication.

A combinatorial approach for small and strong formulations of disjunctive constraints

Joey Huchette

Operations Research Center, MIT, huchette@mit.edu

Juan Pablo Vielma

Sloan School of Management, MIT, jvielma@mit.edu

We present a framework for constructing strong mixed-integer programming formulations for logical disjunctive constraints. Our approach is a generalization of the logarithmically-sized formulations of Vielma and Nemhauser for SOS2 constraints [57], and we offer a complete characterization of its expressive power. We apply the framework to a variety of disjunctive constraints, producing novel small and strong formulations for outer approximations of multilinear terms, generalizations of special ordered sets, piecewise linear functions over a variety of domains, and obstacle avoidance constraints.

MSC2000 subject classification: 90C11

OR/MS subject classification: Primary: Integer Programming; secondary: Piecewise Linear, Polyhedra

History:

1. Introduction A central modeling primitive in mathematical optimization is the disjunctive constraint: any feasible solution must satisfy at least one of some fixed, finite collection of alternatives. This type of constraint is general enough to capture structures as diverse as boolean satisfiability, complementarity constraints, special ordered sets, and (bounded) integrality. The special case of polyhedral disjunctive constraints corresponds to the form

$$x \in \bigcup_{i=1}^d P^i, \quad (1)$$

where we have that each $P^i \subseteq \mathbb{R}^n$ is a polyhedron. In this work, we will focus on \mathcal{V} -polyhedra; that is, we have a description of the P^i in terms of their extreme points $\text{ext}(P^i)$.

We are particularly interested in the case where constraint (1) is *primitive*, or a basic building block for a much more complex optimization problem. For this reason, we are interested in modeling it in a generic, composable way. In particular, if (1) is embedded in a larger, more complex optimization problem

$$\min_{(x,y) \in Q:(1)} f(x,y), \quad (2)$$

we hope for a mathematical description sufficiently structured such that we may use more advanced algorithmic approaches, beyond naïve enumeration, to solve (2). In particular, Q could be described by any number of different types of constraints: linear inequalities, conic constraints, integrality conditions, or additional disjunctive constraints. In this context, it is well known that merely constructing the convex hull $\text{Conv}(\bigcup_{i=1}^d P^i)$ is not sufficient for solving (2); we will need a formulation for $\bigcup_{i=1}^d P^i$ directly.

Mixed-integer programming (MIP) has emerged as an incredibly expressive modeling methodology, with advanced computational methods capable of solving many problems of practical interest, often at very large scale [13, 30]. Constraint (1) with polyhedral sets P^i is particularly well-suited for a mixed-integer programming approach. Indeed, standard formulations for (1) were presented in [29], and are *ideal*, or as strong as possible with respect to their continuous linear programming relaxations (see Section 4.1 for a formal definition). However, this formulation requires introducing d auxiliary binary variables, which may be impractically large, especially in the context of the larger problem (2).

However, it is sometimes possible to construct ideal formulations with considerably fewer auxiliary variables. In particular, a string of recent work [3, 42, 53, 57] has presented ideal formulations for certain highly structured constraints such as SOS2 [9] with only $\mathcal{O}(\log(d))$ auxiliary binary variables and additional constraints (excluding variable bounds). Moreover, these formulations have proven practically useful, and indeed are the most performant by a significant margin for a large swath of instances of the problem classes to which they have been applied [55]. However, building these formulations is complex and ad-hoc, hindering the construction, analysis, and implementation of such formulations for new constraints.

One relatively generic and versatile approach to construct small ideal formulations is the *independent branching* (IB) scheme framework introduced by Vielma and Nemhauser [57]. The approach is to find some (particularly structured) polyhedra $Q^{1,j}$ and $Q^{2,j}$ such that (1) can be rewritten as

$$\bigcup_{i=1}^d P^i = \bigcap_{j=1}^t (Q^{1,j} \cup Q^{2,j}). \quad (3)$$

This represents the disjunctive constraint in term of a series of simple choices between two alternatives. Given such a representation, it is often straightforward to construct a simple, small, and ideal formulation for (1) by formulating each of the t alternatives separately, and then combining them. Furthermore, when the polyhedra P^i are \mathcal{V} -polyhedra, the construction of the independent branching scheme-based formulation is purely combinatorial, based on the extreme points that are shared between the different polyhedra P^i . As we will see, we can therefore approach formulating (1) combinatorially, by studying the shared structure amongst the extreme points.

In this work we generalize and provide a systematic study of the applicability and limitations of the independent branching approach. The contributions of this work can be categorized in the following way.

1. We generalize the notion of independent branching schemes to allow for multiple alternatives, and provide an exact characterization of when there exists *any* independent branching representation for (1), in terms of the graphical representation of the shared extreme points amongst the polyhedra P^i . In particular:

- (a) We demonstrate that the widely-used cardinality constraints cannot be expressed by any independent branching scheme with few alternatives. We argue that this negative result provides theoretical justification for the practical observation that both MIP formulations and simple constraint branching schemes struggle with modeling cardinality constraints effectively.

- (b) We show that arbitrary piecewise linear functions in the plane can be modeled with at most three alternatives, and provide a polynomial-time verifiable condition for representability with two alternatives.

- (c) We argue that nonconvex polygonal set avoidance constraints are always representable with two alternatives.

2. We provide an exact characterization for when there exists a two-alternative independent branching representation for (1) of size t , in terms of the classical biclique covering problem. This relation allows the *algorithmic* construction of small independent branching formulations for (1). In particular, we study and apply simple properties of biclique covers and their composition to systematically construct explicit descriptions of small independent branching formulations for special structures.

3. We apply our framework to a variety of constraints of the form (1) to give an indication of the expressive power of the IB scheme approach and the advantage of using biclique cover techniques to construct formulations. In particular:

(a) We review and develop several generic properties of biclique covers that lead to systematic construction techniques. Using these techniques we construct explicit, small (i.e. logarithmic in d), ideal formulations for generalizations of the special ordered sets of Beale and Tomlin [9], piecewise linear functions over arbitrary 2-dimensional grid triangulations, and outer-approximating discretizations of multilinear terms. This last formulation generalizes the popular logarithmically-sized formulation of Misener et al. [40, 41] for nonconvex quadratic optimization, which we show is not ideal in general.

(b) We provide matching lower bounds for these constructions, showing that they are asymptotically optimal with respect to the size of *any* possible MIP formulation.

2. Preliminaries: Definitions, notation, and nomenclature A (bounded) \mathcal{V} -polyhedra (or polyhedra in \mathcal{V} -form) is a set $P \subset \mathbb{R}^n$ that can be expressed as

$$P = \text{Conv}(V) \stackrel{\text{def}}{=} \left\{ \sum_{v \in V} \lambda_v v : \lambda \in \Delta^V \right\}$$

for some finite set of vectors $V \subset \mathbb{R}^n$, where $\Delta^V \stackrel{\text{def}}{=} \{\lambda \in \mathbb{R}_+^V : \sum_{v \in V} \lambda_v = 1\}$ is the standard simplex. According to the celebrated Minkowski-Weyl Theorem (e.g. [15, Corollary 3.14]), any polyhedral disjunctive constraint (1) can be expressed as the union of \mathcal{V} -polyhedra in terms of their extreme points $\text{ext}(P^i)$ ¹.

When the disjunctive constraint (1) is a union of \mathcal{V} -polyhedra, it suffices to consider only the combinatorial structure of the extreme points of the polyhedra P^i . To see why, consider $J = \bigcup_{i=1}^d \text{ext}(P^i)$ as the *ground set* and take $\mathcal{S} = \{\text{ext}(P^i)\}_{i=1}^d \subseteq 2^J$ as the collection of extreme points for each of the polyhedra. We can then define a corresponding disjunctive constraint that is purely combinatorial on the sets \mathcal{S} .

DEFINITION 1. A *combinatorial disjunctive constraint* (CDC) induced by the sets \mathcal{S} is

$$\lambda \in \text{CDC}(\mathcal{S}) \stackrel{\text{def}}{=} \bigcup_{S \in \mathcal{S}} Q(S),$$

where $Q(S) \stackrel{\text{def}}{=} \{\lambda \in \Delta^J : \lambda_{J \setminus S} \leq 0\}$ is the face that $S \subseteq J$ induces on the standard simplex.

Combinatorial disjunctive constraints may also appear as natural primitive constraints that do not explicitly arise from unions of \mathcal{V} -polyhedra, as we will see in Sections 3.2 and 3.3. However, when they do arise from unions of \mathcal{V} -polyhedra, it is straightforward to construct a corresponding formulation for (1) as

$$\left\{ \sum_{v \in J} \lambda_v v : \lambda \in \text{CDC}(\mathcal{S}) \right\}. \quad (4)$$

One advantage of this approach is that formulation (4) allows us to divorce the problem-specific data (i.e. the values $v \in J$) from the underlying combinatorial structure encapsulated in $\text{CDC}(\mathcal{S})$. As such, we can construct a single, strong formulation for a given structure $\text{CDC}(\mathcal{S})$ and this formulation will remain valid for transformations of the data, so long as this transformation sufficiently preserves the combinatorial structure of $\text{CDC}(\mathcal{S})$. For instance, if $\{P^i\}_{i=1}^d$ are the polyhedra for the original constraint represented by $\text{CDC}(\mathcal{S})$, and $\{\hat{P}^i\}_{i=1}^d$ are those associated with the new data,

¹For the moment we are assuming that the P^i are bounded; the unbounded case is more delicate, as we will discuss shortly.

then a sufficient condition for the formulation of $\text{CDC}(\mathcal{S})$ yielding a valid formulation for $\bigcup_{i=1}^d \hat{P}^i$ is the existence of a bijection $\pi : J \rightarrow \hat{J}$ (with $J = \bigcup_{i=1}^d \text{ext}(P^i)$ and $\hat{J} = \bigcup_{i=1}^d \text{ext}(\hat{P}^i)$) such that

$$v \in \text{ext}(P^i) \iff \pi(v) \in \text{ext}(\hat{P}^i) \quad \forall i \in \llbracket d \rrbracket, v \in J, \quad (5)$$

where $\llbracket d \rrbracket \stackrel{\text{def}}{=} \{1, \dots, d\}$. In this way, we can construct a single small, strong formulation for $\text{CDC}(\mathcal{S})$, and use it repeatedly for many different “combinatorially equivalent” instances of the same constraint.

We note that one subtle disadvantage of this data-agnostic approach is that, even if condition (5) is satisfied, the resulting formulation for $\bigcup_{i=1}^d \hat{P}^i$ may be larger than necessary. An extreme manifestation of this would be when the new polyhedra $\{\hat{P}^i\}_{i=1}^d$ are such that $\hat{P}^i \subseteq \hat{P}^1$ for all $i \in \llbracket d \rrbracket$. In this case, $\bigcup_{i=1}^d \hat{P}^i = \hat{P}^1$, and so formulating this does not require a MIP formulation at all. Less pathological cases could occur where some subset of the disjunctive sets become redundant after changing the problem data. However, we note that in many of the applications considered in this work, the combinatorial representation leads to redundancy of this form only in rare pathological cases (e.g. Sections 3.1 and 3.5). In the remaining cases we will take care to consider, for example, the geometric structure of the data before constructing the disjunctive constraint (e.g. Section 6.2).

Finally, we also note that if we wish to model the case where the polyhedra P^i are unbounded, a result of Jeroslow and Lowe [29] [52, Proposition 11.2] tells us that we may only construct a (binary) MIP formulation for (1) if the recession cones coincide for each P^i . In the case this condition is met, we may formulate (1) with

$$\left\{ \sum_{v \in J} \lambda_v v + \sum_{r \in R} \mu_r r : \lambda \in \text{CDC}(\mathcal{S}), \mu \in \mathbb{R}_+^R \right\}, \quad (6)$$

where R is the shared set of extreme rays for each of the P^i . Therefore, we will restrict our attention to the case where each of the P^i are bounded, as formulating the unbounded case is a straightforward extension.

In the remainder of the paper, we will make the following assumptions on \mathcal{S} that are without loss of generality.

ASSUMPTION 1. *We assume the following about \mathcal{S} .*

- \mathcal{S} is *irredundant*: there do not exist distinct $S, T \in \mathcal{S}$ such that $S \subseteq T$.
- \mathcal{S} *covers the ground set*: $\bigcup_{S \in \mathcal{S}} S = J$.

We will say that a set $S \subseteq J$ is a *feasible set* with respect to $\text{CDC}(\mathcal{S})$ if $Q(S) \subseteq \text{CDC}(\mathcal{S})$ (equivalently, if $S \subseteq T$ for some $T \in \mathcal{S}$) and that it is an *infeasible set* otherwise.

3. Motivating examples As mentioned, we can represent any polyhedral disjunctive constraint (1) as the union of \mathcal{V} -polyhedra. However, there are many disjunctive constraints for which the \mathcal{V} -form of (1) is especially natural. We now present some as running examples that we will return to throughout.

One motif to appear repeatedly will be the graph of a continuous piecewise linear function. That is, given some bounded domain $\Omega \subset \mathbb{R}^n$ and some polyhedral partition $\bigcup_{i=1}^d P^i = \Omega$ (i.e. the relative interiors do not overlap²), we are interested in modeling a continuous function $f : \Omega \rightarrow \mathbb{R}$ such that $x \in P^i \implies f(x) = a^i \cdot x + b_i$ for some appropriate $a^i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$. In order to model the graph $\text{gr}(f; \Omega) \stackrel{\text{def}}{=} \{(x, f(x)) : x \in \Omega\}$, we can construct a formulation for $\text{CDC}(\mathcal{S})$ (with $\mathcal{S} = \{\text{ext}(P^i)\}_{i=1}^d$) and express

$$\text{gr}(f) = \left\{ \sum_{v \in J} \lambda_v (v, f(v)) : \lambda \in \text{CDC}(\mathcal{S}) \right\}, \quad (7)$$

² Formally, $\text{relint}(P^i) \cap \text{relint}(P^j) = \emptyset$ for each $i \neq j$.

where we will use the notation $\text{gr}(f) \equiv \text{gr}(f; \Omega)$ when Ω is clear from context.³

3.1. Univariate piecewise linear functions and the SOS2 constraint Consider a univariate (nonconvex) piecewise linear function f characterized by N breakpoints $x^1 < x^2 < \dots < x^N$. We may model the graph of this function via (7), where $\mathcal{S} = \{\{x^j, x^{j+1}\} : j \in \llbracket N-1 \rrbracket\}$.

As long as the ordering of the breakpoints is preserved, it is easy to see that condition (5) will be satisfied for any transformations of the problem data. Furthermore, the only case in which knowledge of the specific data $\{(x^j, f(x^j))\}_{j=1}^N$ allows the simplification of the original disjunctive representation of $\text{gr}(f)$ is when f is affine in on two adjacent intervals, e.g. affine over $[x^j, x^{j+2}]$ for some $j \in \llbracket N-2 \rrbracket$. Therefore, the potential disadvantage of disregarding the specific data when formulating the constraint occurs only in rare pathological cases which are easy to detect.

For this reason, we will strip out the problem data and instead express the constraint with respect to the indices j of the vertices $\{x^j\}_{j=1}^N$. That is, we take $J = \llbracket N \rrbracket$ and write $\mathcal{S} = \{\{\tau, \tau+1\} : \tau \in \llbracket N-1 \rrbracket\}$ to emphasize the data independence of the constraint and highlight the combinatorial structure. In this form, we can recognize the special ordered set of type 2 (SOS2(N)) constraint of Beale and Tomlin [9], which requires that at most two components of λ may be nonzero, and that these nonzero components must be consecutive in the ordering on J .

3.2. SOS k A generalization of the special ordered sets considers the case where at most k consecutive components of λ may be nonzero at once. In particular, if $J = \llbracket N \rrbracket$, we have $\mathcal{S} = \{\{\tau, \tau+1, \dots, \tau+k-1\} : \tau \in \llbracket N-k+1 \rrbracket\}$. This constraint may arise, for example, in chemical process scheduling problems, where an activated machine may only be on for k consecutive time units and must produce a fixed quantity during that period [21, 33].

3.3. Cardinality constraints An extremely common constraint in optimization is the cardinality constraint of degree ℓ , where at most ℓ components of λ may be nonzero. This corresponds to $\mathcal{S} = \{I \subseteq J : |I| = \ell\}$. A particularly compelling application of the cardinality constraint is in portfolio optimization [11, 12, 14, 54], where it is often advantageous to limit the number of investments to some fixed number ℓ to minimize transaction costs, or to allow differentiation from the performance of the market as a whole.

3.4. Discretizations of multilinear terms Consider a multilinear function $f(x_1, \dots, x_\eta) = \prod_{i=1}^\eta x_i$ defined over some box domain $\Omega \stackrel{\text{def}}{=} [l, u] \subset \mathbb{R}^\eta$. This function appears often in optimization models [22], but is nonconvex, and often leads to problems which are difficult to solve to global optimality in practice [5, 47, 58]. As a result, computational techniques will often “relax” the graph of the function $\text{gr}(f)$ with a convex outer approximation, which is easier to optimize over [49].⁴

For the bilinear case ($\eta = 2$), the well-known McCormick envelope [38] describes the convex hull of $\text{gr}(f)$. Although traditionally stated in an inequality description, we may equivalently describe the convex hull via its four extreme points, which are readily available in closed form. For higher-dimensional multilinear terms, the convex hull has 2^η extreme points, and can be constructed in a similar manner (e.g. see equation (3) in [36] and the associated references).

Misener et al. [40, 41] propose a computational technique for optimizing problems with bilinear terms where, instead of modeling the graph over a single region $\Omega = [l, u] \subset \mathbb{R}^2$, they discretize the region in a regular fashion and apply the McCormick envelope to each subregion. They model this constraint as a union of polyhedra, where each subregion enjoys a tighter relaxation of the bilinear

³We note that the results to follow can potentially be extended to certain discontinuous piecewise linear functions by working instead with the epigraph of f ; we point the interested reader to [55, 56] for further discussion.

⁴Note that these relaxations are useful in the context of global optimization, coupled with algorithmic techniques such as spatial branch-and-bound.

term. Additionally, they propose a logarithmically-sized formulation for the union. However, it is not ideal (see Appendix A), it only applies for bilinear terms ($\eta = 2$), and it is specialized for a particular type of discretization (namely, only discretizing along one component x_1 , and with constant discretization widths $\{h_{j+1}^1 - h_j^1\}_{j=1}^{d_1-1}$).

For a more general setting, we have that the extreme points of the convex hull of the graph $\text{Conv}(\text{gr}(f))$ are given by $\{(x, f(x)) : x \in \text{ext}(\Omega)\}$ [36, equation (3)], where it is easy to see that $\text{ext}(\Omega) = \prod_{i=1}^{\eta} \{l_i, u_i\}$. Consider a grid imposed on $[l, u] \subset \mathbb{R}^{\eta}$; that is, along each component $i \in [\eta]$, we partition $[l_i, u_i]$ along the points $l_i \equiv h_1^i < h_2^i < \dots < h_{d_i-1}^i < h_{d_i}^i \equiv u_i$. This yields $\prod_{i=1}^{\eta} (d_i - 1)$ subregions; denote them by $\mathcal{R} \stackrel{\text{def}}{=} \{\prod_{i=1}^{\eta} [h_{k_i}^i, h_{k_i+1}^i] : k \in \prod_{i=1}^{\eta} \llbracket d_i - 1 \rrbracket\}$.

We can then take the polyhedral partition of Ω given by $P^R \stackrel{\text{def}}{=} \text{Conv}(\text{gr}(f; R))$ for each subregion R , the sets as $\mathcal{S} = \{\text{ext}(P^R)\}_{R \in \mathcal{R}}$, and the ground set as $J = \bigcup \{S \in \mathcal{S}\}$. In particular, we have that $J = \prod_{i=1}^{\eta} \{h_1^i, \dots, h_{d_i}^i\}$. Analogously to the notational simplification we took with the SOS2 constraint, for the remainder we will take $J = \prod_{i=1}^{\eta} \llbracket d_i \rrbracket$ and $\mathcal{S} = \{\prod_{i=1}^{\eta} \{k_i, k_i + 1\} : k \in \prod_{i=1}^{\eta} \llbracket d_i - 1 \rrbracket\}$. We also note that condition (5) is satisfied as long as the ordering of the discretization is respected along each dimension.

3.5. Piecewise linear functions in the plane and grid triangulations Consider a (potentially nonconvex) region $\Omega \subset \mathbb{R}^2$. We would like to model a (also potentially nonconvex) piecewise linear function f with domain over Ω . Take $\{P^i\}_{i=1}^d$ as the set of pieces of the domain, and the corresponding ground set as $J = \bigcup_{i=1}^d \text{ext}(P^i)$ and sets as $\mathcal{S} = \{\text{ext}(P^i)\}_{i=1}^d$. We may then model the piecewise linear function via the graph representation (7).

An important special case occurs when the function f is affine over a triangulation of a grid similar to the one used for multilinear terms. A description of the associated CDC that emphasizes its combinatorial structure can be obtained through the same simplification from Section 3.4 as follows. Consider a rectangular region in the plane $\Omega = [1, M] \times [1, N]$, and the regular grid points $J = \{1, \dots, M\} \times \{1, \dots, N\}$. A *grid triangulation* \mathcal{S} of Ω is then a set \mathcal{S} where:

- Each $S \in \mathcal{S}$ is a triangle: $|S| = 3$.
- \mathcal{S} partitions Ω : $\bigcup_{S \in \mathcal{S}} \text{Conv}(S) = \Omega$ and $\text{relint}(\text{Conv}(S)) \cap \text{relint}(\text{Conv}(T)) = \emptyset$ for each distinct $S, T \in \mathcal{S}$.
- \mathcal{S} is on a regular grid: $S \subset J$ for each $S \in \mathcal{S}$, and $\|v - w\|_{\infty} \leq 1$ for each $v, w \in S$.

As concrete examples, consider Figure 1, where we depict three different triangulations with $M = N = 3$. Grid triangulations are often used to model bivariate, non-separable, piecewise linear functions using (7) [55, 57]. Furthermore, any formulation constructed for a given grid triangulation can be readily applied to any other grid triangulation obtained by shifting the grid points in the plane, so long as the resulting triangulation is strongly isomorphic to, or compatible with, the original triangulation [4].

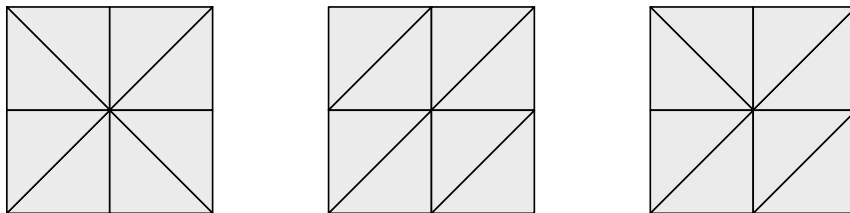


FIGURE 1. Three grid triangulations of $\Omega = [1, 3] \times [1, 3]$: the Union Jack (J1) [51] (Left), the K1 [34] (Center), and a more idiosyncratic construction (Right).

3.6. Obstacle avoidance Consider an unmanned aerial vehicle (UAV) which you would like to navigate through an area with fixed obstacles. At any given time, you wish to impose the constraint that the location of the vehicle $x \in \mathbb{R}^2$ must lie in some (nonconvex) region $\Omega \subset \mathbb{R}^2$, which is the plane, less any obstacles in the area. MIP formulations of this constraint has received interest as a useful primitive for path planning [10, 19, 46, 39].

We may model $x \in \Omega$ by partitioning the region Ω with polyhedra such that $\Omega = \bigcup_{i=1}^d P^i$. Traditional approaches to modeling constraints (1) of this form use a linear inequality description for each of the polyhedra P^i and construct a corresponding big- M formulation [45, 46], which will not be ideal in general. In the \mathcal{V} -polyhedra framework, we will instead take the P^i as \mathcal{V} -polyhedra that partition Ω , and be able to construct small, ideal formulations.

4. MIP formulations for combinatorial disjunctive constraints Using standard MIP formulation techniques, we now present formulations for $\text{CDC}(\mathcal{S})$ for comparison with our approach. In particular, we will argue that the framework we will present later can lead to formulations that are smaller (in terms of the number of auxiliary variables), and enjoy other favorable properties we enumerate in Section 5.

4.1. MIP formulations: Definitions, size, and strength Formally, we say that a (binary) MIP formulation F for a constraint $x \in Q \subseteq \mathbb{R}^{n_1}$ is the composition of linear inequalities (the *linear programming (LP) relaxation*, or just *relaxation* in the context of this work)

$$R \stackrel{\text{def}}{=} \{(x, y, z) \in [l^x, u^x] \times [l^y, u^y] \times [0, 1]^{n_3} : Ax + By + Cz \leq d\}$$

with (binary) integrality conditions

$$F \stackrel{\text{def}}{=} R \cap (\mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \{0, 1\}^{n_3}) \tag{8}$$

such that $\text{Proj}_x(F) = Q$.⁵ Finally, we assume R is line-free (i.e. has at least one extreme point), which is satisfied by essentially all practical formulations.

Throughout, we will be interested in ways of understanding both the strength of a given formulation, as well as in quantifying the size or complexity of a formulation. We say that a formulation is *ideal* if each extreme point of the relaxation naturally satisfies the integrality conditions, i.e. $\text{Proj}_z(\text{ext}(R)) \subseteq \{0, 1\}^{n_3}$. The choice of name is apt, as this is the strongest possible MIP formulation we can expect.⁶

As a measure of the complexity of the formulation, we count the number of auxiliary continuous variables y and continuous binary variables z used by the formulation, as well as the number of inequalities in our description of R . We ignore the size of x , since this is intrinsic to the constraint we wish to model. We will say that a formulation is *extended* if there are auxiliary continuous variables y in the representation F (that is, $n_2 > 0$) and *non-extended* otherwise ($n_2 = 0$). Furthermore, as suggested by the definition of R , we distinguish between variable bounds (e.g. $l^y \leq y \leq u^y$) and general inequalities ($Ax + By + Cz \leq d$), as modern MIP solvers are able to incorporate variable bounds with minimal extra computational cost. Finally, we note that from now on (with a single exception), any statements regarding formulation properties or size will be with respect to the combinatorial disjunctive constraint $\text{CDC}(\mathcal{S})$, rather than the \mathcal{V} -polyhedral disjunctive constraints itself. In other words, all statements are with respect a fixed combinatorial structure, with no possible simplifications from the specific data as discussed in Section 2. The sole exception will be in Section 6.2, where we will make use of some geometric properties inherent in a given realization of the data.

⁵ To handle the unbounded case, we allow the variable bounds $l^x \leq x \leq u^x$ and $l^y \leq y \leq u^y$ to take infinite values.

⁶ An ideal formulation is also *sharp*, i.e. its relaxation projects down to the convex hull of the set we are formulating ($\text{Proj}_x(R) = \text{Conv}(Q)$).

4.2. Existing formulations for combinatorial disjunctive constraints A standard formulation for $\text{CDC}(\mathcal{S})$ adapted from Jeroslow and Lowe [29] is

$$\lambda_v = \sum_{S \in \mathcal{S}: v \in S} \gamma_v^S \quad \forall v \in J \quad (9a)$$

$$z_S = \sum_{v \in S} \gamma_v^S \quad \forall S \in \mathcal{S} \quad (9b)$$

$$\sum_{S \in \mathcal{S}} z_S = 1 \quad (9c)$$

$$\gamma^S \in \Delta^S \quad \forall S \in \mathcal{S} \quad (9d)$$

$$(\lambda, z) \in \Delta^J \times \{0, 1\}^{\mathcal{S}}. \quad (9e)$$

This formulation has $d = |\mathcal{S}|$ auxiliary binary variables, $\sum_{S \in \mathcal{S}} |S|$ auxiliary continuous variables, and no general inequalities. Additionally, it is ideal.

Using Proposition 9.3 from [52], we can construct an ideal MIP formulation with fewer auxiliary binary variables:

$$\lambda_v = \sum_{S \in \mathcal{S}: v \in S} \gamma_v^S \quad \forall v \in J \quad (10a)$$

$$\sum_{S \in \mathcal{S}} \sum_{v \in S} \gamma_v^S = 1 \quad (10b)$$

$$\sum_{S \in \mathcal{S}} \sum_{v \in S} h^S \gamma_v^S = z \quad (10c)$$

$$\gamma^S \geq 0 \quad \forall S \in \mathcal{S} \quad (10d)$$

$$z \in \{0, 1\}^r, \quad (10e)$$

where $\{h^S\}_{S \in \mathcal{S}} \subseteq \{0, 1\}^r$ is some set of distinct binary vectors. This formulation is actually a generalization of (9), which we recover if we take $h^S = \mathbf{e}^S \in \mathbb{R}^S$ as the canonical unit vectors. If instead we take r to be as small as possible (while ensuring that the vectors $\{h^S\}_{S \in \mathcal{S}}$ are distinct), we recover $r = \lceil \log_2(d) \rceil$. Therefore, formulation (10) yields an ideal extended formulation for (1) with $\lceil \log_2(d) \rceil$ auxiliary binary variables, $\sum_{S \in \mathcal{S}} |S|$ auxiliary continuous variables, and no general inequalities. The following corollary shows that this is the smallest number of auxiliary binary variables we may hope for.

PROPOSITION 1. *If the sets \mathcal{S} are irredundant, then any binary MIP formulation for $\text{CDC}(\mathcal{S})$ must have at least $\lceil \log_2(d) \rceil$ auxiliary binary variables.*

Proof See Appendix B. \square

The formulations thus far have been extended formulations, as they are constructed by formulating each polyhedra $Q(S)$ separately and then aggregating them, rather than working with the combinatorial structure underlying the shared extreme points. Therefore, each of these formulations requires a copy of the multiplier γ_v^S for each set $S \in \mathcal{S}$ for which $v \in S$, and so $\sum_{i=1}^d |\mathcal{S}|$ auxiliary continuous variables total.

In contrast, we can construct non-extended formulations for $\text{CDC}(\mathcal{S})$ that work directly on the λ variables and the underlying combinatorial structure of \mathcal{S} . An example of a non-extended formulation for CDC is the widely used ad-hoc formulation (see [52, Section 6] and the references therein) given by

$$\lambda_v \leq \sum_{S \in \mathcal{S}: v \in S} z_S \quad \forall v \in J \quad (11a)$$

$$\sum_{S \in \mathcal{S}} z_S = 1 \quad (11b)$$

$$(\lambda, z) \in \Delta^J \times \{0, 1\}^{\mathcal{S}}. \quad (11c)$$

This formulation is not necessarily ideal, and it requires no auxiliary continuous variables, d auxiliary binary variables, and $|J|$ general inequalities.

In summary, we have seen an ideal extended formulation (10) for $\text{CDC}(\mathcal{S})$ with relatively few auxiliary binary variables, but relatively many auxiliary continuous variables. On the other end of the spectrum, we have a non-extended formulation (11) with no auxiliary continuous variables, but which requires relatively many auxiliary binary variables and which may fail to be ideal. However, we know that in special cases we can construct ideal, non-extended formulations with only $\mathcal{O}(\log(d))$ auxiliary variables and constraints (e.g. SOS1, SOS2, and particular 2-dimensional grid triangulations [53, 57]). This work provides a framework for constructing such *small, strong*, non-extended MIP formulations for $\text{CDC}(\mathcal{S})$, which are automatically ideal, and in the best case will have $\mathcal{O}(\log(d))$ auxiliary binary variables and general inequality constraints.

5. Independent branching schemes Vielma and Nemhauser [57] introduced the notion of an *independent branching scheme* as a natural framework for constructing formulations for combinatorial disjunctive constraints. The independent branching scheme is a logically equivalent way of expressing a CDC in terms of a conjunction of dichotomies: that is, as a series of choices between two (simple) options. This approach is parsimonious: if you are given an independent branching scheme for a particular CDC, it is straightforward to construct an ideal formulation whose size is on the order of the number of dichotomies. For our purposes, we present a generalized notion, where we allow potentially more than two alternatives.

DEFINITION 2. A k -way independent branching (IB) scheme of depth t for $\text{CDC}(\mathcal{S})$ is given by a family of sets (L_1^j, \dots, L_k^j) (where each $L_i^j \subseteq J$) for $j \in \llbracket t \rrbracket$, where

$$\text{CDC}(\mathcal{S}) = \bigcap_{j=1}^t \left(\bigcup_{i=1}^k Q(L_i^j) \right). \quad (12)$$

We say that such an IB scheme has *depth* t , and that each $j \in \llbracket t \rrbracket$ yields a corresponding *level* of the IB scheme $\bigcup_{i=1}^k Q(L_i^j)$, given by the k *alternatives* $Q(L_i^j)$.

An equivalent way of understanding these representations, which we will be using for the remainder of this work, is by eschewing the polyhedra $Q(L_i^j)$ and working directly on the underlying set L_i^j . That is, a valid k -way IB scheme satisfies the condition that

$$T \subseteq J \text{ is a feasible set} \iff \forall j \in \llbracket t \rrbracket, \exists i \in \llbracket k \rrbracket \text{ s.t. } T \subseteq L_i^j.$$

First, we observe that, due to our assumption that \mathcal{S} covers the ground set, we have that for each element $v \in J$ and level $j \in \llbracket t \rrbracket$, there will be at least one alternative $i \in \llbracket k \rrbracket$ such that $v \in L_i^j$. We will use this extensively in the analysis to come, as it simplifies some otherwise tedious case analyses. Second, we see that this definition can capture potential schemes with a variable number of alternatives in each level by adding empty alternatives $L_i^j = \emptyset$, provided we take k as the maximum number of alternatives for all levels. For notational simplicity, we say that a 2-way IB scheme is a *pairwise IB scheme*, and in this case we write the sets as $\{(L^j, R^j)\}_{j=1}^t$ as in [57]. In contrast, we will call the case with $k > 2$ a *multi-way IB scheme*.

In this form, we have replaced the monolithic constraint $\text{CDC}(\mathcal{S})$ by t constraints, each of which require the selection between k alternatives. We may then use standard techniques to construct a corresponding mixed-integer formulation.

PROPOSITION 2. *Given an independent branching scheme $\{(L_1^j, \dots, L_k^j)\}_{j=1}^t$ for $\text{CDC}(\mathcal{S})$, the following is a valid formulation for $\text{CDC}(\mathcal{S})$:*

$$\sum_{v \notin L_i^j} \lambda_v \leq 1 - z_i^j \quad \forall j \in \llbracket t \rrbracket, \forall i \in \llbracket k \rrbracket \quad (13a)$$

$$\sum_{i=1}^k z_i^j = 1 \quad \forall j \in \llbracket t \rrbracket \quad (13b)$$

$$\lambda \in \Delta^J \quad (13c)$$

$$z^j \in \{0, 1\}^k \quad \forall j \in \llbracket t \rrbracket. \quad (13d)$$

The formulation is known to be ideal for $k = 2$ [55, 57]. It has no auxiliary continuous variables, kt auxiliary binary variables, and kt general inequalities.

5.1. Constraint branching via independent branching-based formulations The canonical algorithmic technique for solving mixed-integer programming problems is some variation of branch-and-bound [35], which implicitly enumerates all possible values for the binary variables. In its simplest form, a sequence of problems are solved, starting with the relaxation of the MIP formulation, after which a binary variable z_i is chosen for *branching*. That is, the current problem is branched into two subproblems: one with the additional constraint $z_i \leq 0$, another with $z_i \geq 1$. Repeating this procedure, the subproblems form a (binary) tree whose leaves correspond to all 2^{n_3} possible values for the n_3 binary variables in formulation (8). At any given subproblem, the augmented relaxation to be solved is described by the set of binary variables fixed to zero, and the set of those fixed to one.

The spirit of constraint branching is to allow richer branching decisions. For example, a branching decision might be between k alternatives of the form $\{Q^i\}_{i=1}^k$, where each Q^i is formed by adding a general inequality constraint to the existing relaxation at the current node. This concept has significant overlap with the broader field of constraint programming [7, 28], which has been recognized and exploited in the mixed-integer programming literature [1, 6, 25, 44, 48]. More complex constraint branching can often lead to a more balanced branch-and-bound tree, which can significantly improve computational performance (see, for example, [52, Section 8] and [59] for more discussion). Combinatorial disjunctive constraints are a natural setting to apply constraint branching directly on the continuous λ variables [9, 18, 31, 32, 37]. Indeed, the classical examples of the SOS1 and SOS2 constraints [9] show that we do not necessarily require a MIP formulation (or the auxiliary binary variables z) for modeling combinatorial disjunctive constraints, as the disjunction can be enforced directly through constraint branching on the λ variables. These constraint branching approaches without auxiliary binary variables can be implemented in an ad-hoc branch-and-bound procedure, or through branching callbacks available in some MIP solvers, such as CPLEX. In theory, this approach should outperform a MIP formulation like (11) that introduces additional variables and constraints. However, realizing this performance advantage in practice can require significant effort and technical expertise. For instance, Vielma et al. [55, 56] observe that the basic formulation (11) clearly outperformed the SOS2 branching implementation in CPLEX v9.1. However, CPLEX v11 implemented an optimized version of SOS2 branching that used the advanced branch selection techniques available for variable branching, reversing this performance gap with respect the MIP formulation approach.

One way to avoid re-implementing the advanced branching selection techniques for a new constraint branching approach is by constructing a MIP model that automatically inherits the advanced constraint branch selection, but using the solvers traditional variable branching [6, 57]. For simplicity, assume that the constraint branching approach has t branching options, each of which creates k branches, and that each constraint added has support on the λ variables, with variable coefficients in $\{0, 1\}$ and a zero right-hand-side. That is, branch $i \in \llbracket k \rrbracket$ of branching option $j \in \llbracket t \rrbracket$ adds a constraint of the form $\sum_{v \notin L_i^j} \lambda_v \leq 0$. This is equivalent to a *multi-variable* branching approach that fixes groups of variables to zero, and it includes as special cases most constraint branching approaches, including SOS1/SOS2 branching. Then (13) is a MIP formulation for this

multi-variable constraint branching scheme, as variable branching on $\{z_i^j\}_{i=1}^k$ enforces constraint branching option j on the λ variables.

This connection highlights the natural theoretical equivalence between a multi-variable branching and an independent branching formulation. A practical difference between the two is that direct multi-variable branching must implement an explicit branch selection and implementation routine, while an independent branching formulation inherits the variable branching selection and implementation routines of the MIP solver. The upshot of this is that the independent branching formulation must provide a complete catalog of all possible branching options up-front (i.e. through the formulation), while direct multi-variable branching can have a large catalog of branching options that are implicitly defined by the branching routines.

We finally note that, as discussed by Vielma and Nemhauser [57, Section 3], variable branching on a non-IB formulation such as (10) can fix components of λ to zero in a fashion that is *dependent* on previous branching decisions. This is not the case with independent branching formulations, hence the name. We review this independence property in detail in Appendix C.

6. Independent branching scheme representability To start, we observe that the independent branching approach is not sufficiently general to capture every possible formulation for $\text{CDC}(\mathcal{S})$. In particular, there is the restriction that each alternative $Q(L_i^j)$ restricts the λ variables to lie on a *single* face of the standard simplex. A natural first question is then: given a family of sets \mathcal{S} , do any k -way IB schemes exist for $\text{CDC}(\mathcal{S})$? We provide an answer, based on a graphical characterization of the constraint.

DEFINITION 3. Let $H \stackrel{\text{def}}{=} (J, \mathcal{E})$ be a hypergraph with hyperedge set $\mathcal{E} \subseteq 2^J$.

- The *rank* of H is $r(H) \stackrel{\text{def}}{=} \max\{|E| : E \in \mathcal{E}\}$.
- A (*weakly*) *independent set* of H is a set $U \subseteq J$ that does not contain any element of \mathcal{E} as a subset.
- The *conflict hypergraph* of \mathcal{S} is $H_{\mathcal{S}}^c \stackrel{\text{def}}{=} (J, \mathcal{E}_{\mathcal{S}})$, where $\mathcal{E}_{\mathcal{S}} \stackrel{\text{def}}{=} \{E \subseteq J : E \text{ is a minimal infeasible set}\}$.

LEMMA 1. *The maximal independent sets S in $H_{\mathcal{S}}^c$ are exactly the sets $S \in \mathcal{S}$.*

Proof If $S \in \mathcal{S}$, it is obviously a feasible set, and so we have immediately that S is an independent set in $H_{\mathcal{S}}^c$. If it is not maximal, then we could add some $v \in J \setminus S$ and maintain feasibility, which would violate our nonredundancy assumption (i.e. $S \cup \{v\} \subseteq S' \in \mathcal{S}$, $S \in \mathcal{S}$, and $S \subsetneq S' \subseteq S$).

If S is a maximal independent set in $H_{\mathcal{S}}^c$, then it must be a feasible set with respect to \mathcal{S} as well. As it is maximal, there is no set $\hat{S} \in \mathcal{S}$ with $S \subsetneq \hat{S}$, and so we must have $S \in \mathcal{S}$ as well. \square

THEOREM 1. *A k -way IB scheme exists for $\text{CDC}(\mathcal{S})$ if and only if $r(H_{\mathcal{S}}^c) \leq k$. In particular, if $\mathcal{E}_{\mathcal{S}} = \left\{ E^j = \{e_1^j, \dots, e_{|E^j|}^j\} \right\}_{j=1}^t$ is the hyperedge set for the conflict hypergraph $H_{\mathcal{S}}^c$, then an $r(H_{\mathcal{S}}^c)$ -way IB scheme for $\text{CDC}(\mathcal{S})$ is given by*

$$L_i^j = \begin{cases} J \setminus \{e_i^j\} & i \leq |E^j| \\ \emptyset & \text{o.w.} \end{cases} \quad \forall i \in \llbracket r(H_{\mathcal{S}}^c) \rrbracket, j \in \llbracket t \rrbracket. \quad (14)$$

Proof To show the “if” direction, it suffices to show the validity of (14). First note that every minimally infeasible set $E^j \in \mathcal{E}_{\mathcal{S}}$ is rendered infeasible by level j , which implies that every infeasible set is rendered infeasible as well. Then note that for any $S \in \mathcal{S}$ and for any $j \in \llbracket t \rrbracket$, we have $E^j \not\subseteq S$, so there exists $i \in \llbracket |E^j| \rrbracket$ such that $e_i^j \in E^j \setminus S$. Hence, $S \in L_i^j$ and S is feasible for level j .

To show the “only if” direction, assume for a contradiction that there exists a k -way IB scheme with $k \leq r(H_{\mathcal{S}}^c) - 1$. Take a minimal infeasible set $E = \{e_1, \dots, e_r\} \in \mathcal{E}_{\mathcal{S}}$, where $r = r(H_{\mathcal{S}}^c)$. Then take $j \in \llbracket t \rrbracket$ as a level of the IB scheme that renders E infeasible. By the minimality of E , we have that, for all $\ell \in \llbracket r \rrbracket$, there exists some $i(\ell) \in \llbracket k \rrbracket$ such that $E(\ell) \stackrel{\text{def}}{=} E \setminus \{e_{\ell}\} \subseteq L_{i(\ell)}^j$. As $k < r$, we may apply

the pigeonhole principle to see that there must exist some distinct $\ell^1, \ell^2 \in \llbracket r \rrbracket$ such that $i(\ell^1) = i(\ell^2)$, and such that $E(\ell^1) \subseteq L_{i(\ell^1)}^j$ and $E(\ell^2) \subseteq L_{i(\ell^2)}^j$. As $E = E(\ell^1) \cup E(\ell^2)$ and $L_{i(\ell^1)}^j = L_{i(\ell^2)}^j$, this implies that $E \subseteq L_{i(\ell^1)}^j$, which contradicts our supposition that level j rendering E infeasible. \square

Throughout, we will say that $\text{CDC}(\mathcal{S})$ is k -way IB-representable (or pairwise IB-representable for $k = 2$) if it admits a k -way IB scheme.

6.1. Cardinality constraints Our first application of Theorem 1 is to derive a strong restriction on the existence of multi-way IB schemes for the cardinality constraint.

COROLLARY 1. *A cardinality constraint of degree ℓ is k -way IB-representable if and only if $k > \ell$.*

Proof Direct from Theorem 1 by observing that $r(H_{\mathcal{S}}^c) = \ell + 1$. \square

We observe that the IB scheme (14), when applied to the cardinality constraint, is a natural MIP formulation for the ‘‘conjunctive normal form’’ [8], and is unlikely to be practical for even moderately large ℓ . In addition, both specialized constraint branching schemes for cardinality constraints [27] and the binary variable branching induced by standard formulations for cardinality constraints are quite imbalanced. The existence of a pairwise independent branching scheme for cardinality constraints would likely have finally produced the sought-after balanced constraint branching. However, Corollary 1 implies that such a balanced constraint branching cannot be produced via IB schemes, or equivalently by constraint branchings that do not use general inequalities (i.e. are only multi-variable branchings).

6.2. Polygonal partitions of the plane Consider a (nonconvex) bounded region in the plane $\Omega \subset \mathbb{R}^2$ that describes all possible locations for a UAV, as described in Section 3.6. Assume that Ω can be partitioned into polyhedra $\{P^i\}_{i=1}^d$ such that $\bigcup_{i=1}^d P^i = \Omega$ and $\text{relint}(P^i) \cap \text{relint}(P^j) = \emptyset$ for each distinct $i, j \in \llbracket d \rrbracket$. We note that this partition will not, in general, be unique, and its selection can have a significant effect on questions of representability or formulation size. Figure 2 illustrates this for a convex region with a ‘‘hole.’’ The figure shows three ways to partition the resulting nonconvex region into convex polyhedra. Once this partition is fixed, we describe the associated polyhedra in \mathcal{V} -form, and so the corresponding combinatorial disjunctive constraint is given by $\mathcal{S} = \{\text{ext}(P^i)\}_{i=1}^d$ and $J = \bigcup \{S \in \mathcal{S}\}$. We additionally forbid polyhedra with ‘‘internal vertices’’ by requiring that

$$v \in P^i \iff v \in \text{ext}(P^i) \quad \forall i \in \llbracket d \rrbracket, v \in J, \quad (15)$$

so that \mathcal{S} corresponds to the maximal elements of a polyhedral complex [60, Section 5.1]. For example, the second and third partitions in Figure 2 satisfy this condition, while the first does not.

In this setting, minimal infeasible sets have a natural characterization.

THEOREM 2. *Take bounded $\Omega \subset \mathbb{R}^2$ and a polyhedral partition $\{P^i\}_{i=1}^d$ of Ω satisfying the internal vertex condition (15). If $\mathcal{S} = \{\text{ext}(P^i)\}_{i=1}^d$, then $r(H_{\mathcal{S}}^c) \leq 3$.*

Proof Take some minimal infeasible hyperedge $E \in \mathcal{E}_{\mathcal{S}}$ of $H_{\mathcal{S}}^c$, assuming for contradiction that $r = |E| > 3$, and label the points $E = \{v^i\}_{i=1}^r$. First, we show that the points may not be in general position, i.e. that w.l.o.g. $v^r \in \text{Conv}(\{v^i\}_{i=1}^{r-1})$. Then, we argue that the points not being in general position implies that $\{v^i\}_{i=1}^{r-1}$ is also an infeasible set, violating the minimality condition.

Assume for contradiction that the points are in general position; that is, that none can be written as a convex combination of the others. This implies that $\text{ext}(\text{Conv}(E)) = E$. Assume that the ordering $\{v^1, \dots, v^r\}$ forms a path around the edges of $\text{Conv}(E)$; that is, v^i and v^j both lie on an edge of $\text{Conv}(E)$ if and only if $|i - j| = 1$ or $\{i, j\} = \{1, r\}$.

Choose some set $S^1 \in \mathcal{S}$ and some $2 < j < r$ such that $v^1, v^j \in S^1$ and $v^2 \notin S^1$; the associated polyhedron is P^1 . Such a set exists, else E is not a minimal infeasible set (choose instead $E \setminus \{v^2\}$). Now choose $S^2 \in \mathcal{S}$ such that $v^2, v^r \in S^2$; the associated polyhedron is P^2 . Such a set exists, as $\{v^2, v^r\} \subsetneq E$ and E is minimal. As the nodes v^1, v^2, v^j, v^r are interlaced along the boundary of $\text{Conv}(E)$, we have that $\text{Conv}(\{v^1, v^j\}) \cap \text{Conv}(\{v^2, v^r\}) \subset \text{Conv}(E)$ is nonempty. As each of the four points is on the boundary of $\text{Conv}(E)$, and the points are in general position, it follows that $\text{Conv}(\{v^1, v^j\}) \cap \text{Conv}(\{v^2, v^r\}) = \text{relint}(\text{Conv}(\{v^1, v^j\})) \cap \text{relint}(\text{Conv}(\{v^2, v^r\}))$. Therefore, there must exist some point y with $y \in \text{relint}(\text{Conv}(\{v^1, v^j\})) \subseteq \text{relint}(P^1)$ and $y \in \text{relint}(\text{Conv}(\{v^2, v^r\})) \subseteq \text{relint}(P^2)$. However, this implies that $\text{relint}(P^1) \cap \text{relint}(P^2) \neq \emptyset$, which contradicts the assumption that our sets partition the region Ω .

Finally, it just remains to show that $\{v^i\}_{i=1}^{r-1}$ is also an infeasible set, and therefore $\{v^i\}_{i=1}^r$ cannot be a minimal infeasible set. Assume for contradiction that it is not: i.e. that there exists some j such that $\{v^i\}_{i=1}^{r-1} \subseteq \text{ext}(P^j)$. But this implies that $v^r \in J$ and $v^r \in \text{Conv}(E) \subseteq P^j$, yet $v^r \notin \text{ext}(P^j)$, a contradiction of the internal vertices assumption. \square

In other words, every polyhedral partition of the plane is 3-way independent branching-representable, and pairwise IB representability can be checked in time polynomial in $|J|$ (for example, by enumerating the subsets of J of cardinality 3). To illustrate, in Figure 2 we depict the three possible cases for a partition with respect to Theorem 2: 1) it does not satisfy the internal vertices condition, 2) it admits a pairwise IB scheme ($r(H_S^c) = 2$), or 3) it does not admit a pairwise IB scheme, but does admit a 3-way IB scheme ($r(H_S^c) = 3$).

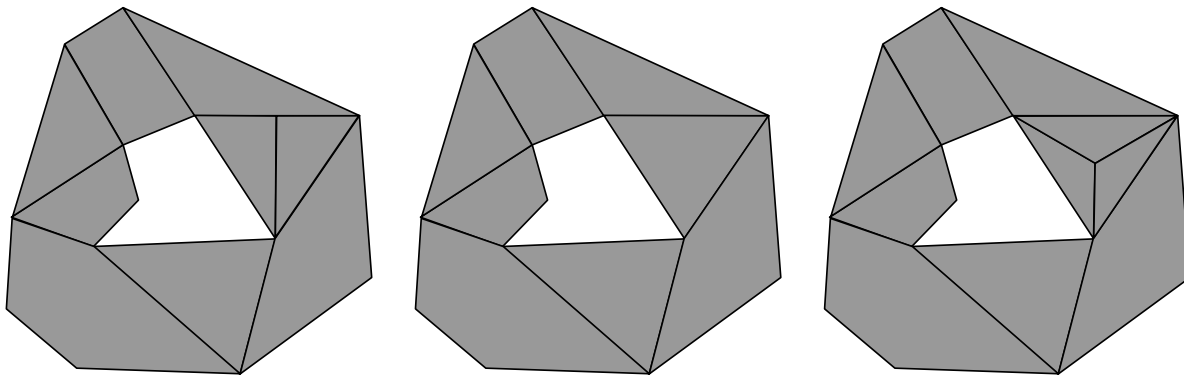


FIGURE 2. Partitions of a nonconvex region in the plane obtained by removing a central non-convex portion from a convex polyhedron. The first partition does not satisfy the internal vertices condition (15) (Left), the second partition admits a pairwise IB scheme (Center), and the third partition admits a 3-way IB scheme but not a pairwise one (Right).

Furthermore, we can argue that we can always represent an obstacle avoidance constraint in such a way that it admits a pairwise IB scheme. Inspecting Figure 2, we see that the region Ω is the same in each, and it is only the partition of Ω that can potentially lead to constraints that are not pairwise IB-representable. Therefore, the obstacle avoidance constraint is invariant to the specification of the partition, and if any polyhedral partitioning exists, then it is always possible to construct one that satisfies the conditions of Theorem 2.⁷

⁷Note that this result does not carry over to piecewise linear functions over Ω , as the choice of the partition is intimately connected with the values the function may take.

6.3. SOS2 Our first example of a constraint that is always pairwise IB-representable is the SOS2(N) constraint. Recall that $J = \llbracket N \rrbracket$ and $\mathcal{S} = \{\{\tau, \tau + 1\} : \tau \in \llbracket N - 1 \rrbracket\}$ for SOS2. Then $\mathcal{E}_\mathcal{S} = \{\{\tau, \tau + t\} : \tau, \tau + t \in \llbracket N \rrbracket, t \geq 2\}$, $r(H_\mathcal{S}^c) = 2$, and formulation (14) has depth $t = \binom{N}{2} - N + 1$. However, Vielma and Nemhauser [57] construct a pairwise IB scheme for SOS2(N) constraints of depth logarithmic in N . The construction is built around a Gray code [50], or sequence of distinct binary vectors $\{h^i\}_{i=1}^{N-1} \subseteq \{0, 1\}^{\lceil \log_2(N-1) \rceil}$ where each adjacent pair (h^i, h^{i+1}) differs in exactly one component. Notationally, here and throughout, take $h^0 \stackrel{\text{def}}{=} h^1$ and $h^N \stackrel{\text{def}}{=} h^{N-1}$. The pairwise IB scheme is then given by

$$L^j = \{\tau \in \llbracket N \rrbracket : h_j^{\tau-1} = 1 \text{ or } h_j^\tau = 1\}, \quad R^j = \{\tau \in \llbracket N \rrbracket : h_j^{\tau-1} = 0 \text{ or } h_j^\tau = 0\} \quad \forall j \in \llbracket \lceil \log_2(N-1) \rceil \rrbracket.$$

We observe that the resulting formulation matches the lower bound from Proposition 1 with respect to the number of auxiliary binary variables and is significantly smaller than formulation (14).

Indeed, formulation (14) is likely to be unnecessarily large for the other constraints we consider as well, so we turn our attention to finding smaller IB schemes in Section 7. We end this section by applying Theorem 1 to succinctly prove the pairwise IB-representability of two other constraints.

6.4. Other pairwise IB-representable constraints

SOS k In this case $\mathcal{E}_\mathcal{S} = \{\{\tau, \tau + t\} : \tau, \tau + t \in \llbracket N \rrbracket, t \geq k + 1\}$ and $r(H_\mathcal{S}^c) = 2$.

Grid triangulations We show that $r(H_\mathcal{S}^c) = 2$ by seeing that for any infeasible set $T \subseteq J$ there exist some distinct $v, w \in T$ such that $\{v, w\}$ is infeasible. Indeed, if there are some $v, w \in T$ such that $\|v - w\|_\infty > 1$, then there does not exist any triangle on the grid that contains both, so $\{v, w\}$ is also an infeasible set. Otherwise, we have that $T \subset \{r, r + 1\} \times \{s, s + 1\}$ for some r, s , and that T contains elements in both of the triangles in this square. For each of the two triangles, we can select an element of T that is not contained in the other triangle, which yields an infeasible pair contained in T . Therefore, any grid triangulation is pairwise IB-representable.

7. Pairwise independent branching schemes The pairwise independent branching scheme framework was initially introduced by Vielma and Nemhauser [57], where it was used to model particularly structured piecewise linear functions. In the remainder of this work, we will focus on pairwise IB schemes and offer a complete picture of their expressive powers, along with an algorithmic framework for constructing them.

7.1. Graphical representations of pairwise IB-representable CDCs From our covering assumption $J = \bigcup \{S \in \mathcal{S}\}$, we can see that $|E| \geq 2$ for each $E \in \mathcal{E}_\mathcal{S}$. By applying Theorem 1, we then immediately have that $\text{CDC}(\mathcal{S})$ is pairwise IB-representable if and only if $H_\mathcal{S}^c$ is (equivalent to) a graph. Along this line, for any constraint we may define a *conflict graph* for any $\text{CDC}(\mathcal{S})$ as $G_\mathcal{S}^c \stackrel{\text{def}}{=} (J, \bar{E})$, where $\bar{E} = \bar{E}_\mathcal{S} \stackrel{\text{def}}{=} \{\{u, v\} \in [J]^2 : u \neq v, \{u, v\} \text{ is an infeasible set}\}$ is the set of all infeasible pairs of elements of J . Checking for pairwise IB-representability then reduces to verifying if $\mathcal{E}_\mathcal{S} = \bar{E}_\mathcal{S}$. The following corollary of Theorem 1 shows that this can also be verified by only working with $G_\mathcal{S}^c$.

COROLLARY 2. *$\text{CDC}(\mathcal{S})$ is pairwise IB-representable if and only if the sets \mathcal{S} are exactly the maximal independent sets of $G_\mathcal{S}^c$.*

Proof If $\text{CDC}(\mathcal{S})$ is pairwise IB-representable, then $G_\mathcal{S}^c$ is equivalent to $H_\mathcal{S}^c$. By applying Theorem 1, the maximal independent sets of $G_\mathcal{S}^c$ are exactly the elements of \mathcal{S} . For the converse, assume for a contradiction that \mathcal{S} is exactly the maximal independent sets of $G_\mathcal{S}^c$, but that there exists some $E \in \mathcal{E}_\mathcal{S}$ with $|E| \geq 3$. By the minimal infeasibility of E , we have that $\{r, s\} \notin \bar{E}_\mathcal{S}$ for any distinct $r, s \in E$, and therefore E is an independent set in $G_\mathcal{S}^c$. This implies that E is contained in a maximal independent S . By assumption, $S \in \mathcal{S}$, which contradicts the infeasibility of E . \square

Therefore, verifying general pairwise IB-representability reduces to enumerating the maximal independent sets of G_S^c and identifying them to exactly the sets \mathcal{S} . As an example, we can see that, for cardinality constraint of degree ℓ with $2 \leq \ell < |J|$, the only maximal independent set of G_S^c is the entire ground set J , which certainly cannot be identified with $\mathcal{S} = \{S \subset J : |S| = \ell\}$.

7.2. Representation at a given depth Once a CDC has been shown to be pairwise IB-representable, a natural next question is: what is the smallest possible depth at which we may construct an IB scheme? That is, we ask if there exists a pairwise IB scheme for $\text{CDC}(\mathcal{S})$ of some given depth t . The answer to this question reduces to the existence of a graphical decomposition of the conflict graph G_S^c .

DEFINITION 4. A *biclique cover* of the graph $G = (J, E)$ is a collection of complete bipartite subgraphs $\left\{ G^j \stackrel{\text{def}}{=} (A^j \cup B^j, E^j) \right\}_{j=1}^t$ of G that cover all the edges of G . Formally, this means that for each $t \in \llbracket t \rrbracket$, $\emptyset \subsetneq A^j, B^j \subsetneq J$, $A^j \cap B^j = \emptyset$, and $E^j = A^j * B^j \stackrel{\text{def}}{=} \{\{a, b\} : a \in A^j, b \in B^j\}$, and that $\bigcup_{j=1}^t E^j = E$. For notational simplicity, we will often refer to the sets $\{(A^j, B^j)\}_{j=1}^t$ as a biclique cover, as we can recover the graphs G^j directly.

The following theorem formalizes the equivalence between biclique covers and pairwise IB schemes.

THEOREM 3. *If $\{(A^j, B^j)\}_{j=1}^t$ is biclique cover of the conflict graph G_S^c for pairwise IB-representable $\text{CDC}(\mathcal{S})$, then a pairwise IB scheme for $\text{CDC}(\mathcal{S})$ is given by*

$$L^j = J \setminus A^j, \quad R^j = J \setminus B^j \quad \forall j \in \llbracket t \rrbracket. \quad (16)$$

Conversely, if $\{(L^j, R^j)\}_{j=1}^t$ is a pairwise IB scheme for $\text{CDC}(\mathcal{S})$, then a biclique cover of the conflict graph G_S^c is given by

$$A^j = J \setminus L^j, \quad B^j = J \setminus R^j \quad \forall j \in \llbracket t \rrbracket. \quad (17)$$

Proof For the first part, take \bar{E} as the edge set of G_S^c . To see that any $S \in \mathcal{S}$ is feasible for the IB scheme (16), note that if $S \not\subseteq L^j$ and $S \not\subseteq R^j$, then there exist some $u \in A^j \cap S$ and $v \in B^j \cap S$. However, this implies that $\{u, v\} \in A^j * B^j \subseteq \bar{E}$, which is a contradiction of feasibility as $\{u, v\} \subseteq S$ and $S \in \mathcal{S}$. Furthermore, as $\{(A^j, B^j)\}_{j=1}^t$ is a biclique cover of G_S^c , for every $\{u, v\} \in \bar{E}$ we have that there exists some level $j \in \llbracket t \rrbracket$ such that w.l.o.g. $u \in A^j$ and $v \in B^j$. This implies that $u \notin L^j$ and $v \notin R^j$ by their construction, and as $\text{CDC}(\mathcal{S})$ is pairwise IB-representable, then any infeasible set for $\text{CDC}(\mathcal{S})$ is also infeasible for the proposed IB scheme. Therefore, (16) is a valid pairwise IB scheme.

For the second part, note that $A^j \cap B^j = \emptyset$ for all $j \in \llbracket t \rrbracket$, and that the covering portion of Assumption 1 implies that $L^j \cup R^j = J$. Therefore, it only remains to show that $\bar{E} = \bigcup_{j=1}^t \bar{E}^j$, where $\bar{E}^j = A^j * B^j$. For that, first note that as $L^j \cup R^j = J$, we have that $A^j = R^j \setminus L^j$ and $B^j = L^j \setminus R^j$. The containment $\bar{E} \subseteq \bigcup_{j=1}^t \bar{E}^j$ then follows by noting that, as $\{(L^j, R^j)\}_{j=1}^t$ is a valid pairwise IB scheme, each minimal infeasible set $\{u, v\} \in \bar{E}$ has some level $j \in \llbracket t \rrbracket$ such that $\{u, v\} \not\subseteq L^j$ and $\{u, v\} \not\subseteq R^j$. Then, as $L^j \cup R^j = J$, we have (w.l.o.g.) that $u \in L^j \setminus R^j \equiv B^j$ and $v \in R^j \setminus L^j \equiv A^j$, and so $\{u, v\} \in \bar{E}^j$. For the reverse containment $\bigcup_{j=1}^t \bar{E}^j \subseteq \bar{E}$, take some arbitrary $j \in \llbracket t \rrbracket$ and some edge $\{a, b\} \in \bar{E}^j$. From the definition of our biclique cover, we have that w.l.o.g. $a \in A^j \equiv R^j \setminus L^j$ and $b \in B^j \equiv L^j \setminus R^j$. Therefore, $\{a, b\}$ is an infeasible set for the IB scheme, and thus for $\text{CDC}(\mathcal{S})$ as well, and so $\{a, b\} \in \bar{E}$. \square

We can now naturally frame the problem of finding a minimum depth pairwise IB scheme as the minimum biclique cover problem [20, 23]. Unfortunately, the decision version of this problem is known to be NP-complete [43] and inapproximable within a factor of $|J|^{1/3-\epsilon}$ if $P \neq NP$ [24], even for bipartite graphs. However, we note that it is simple to construct a MIP feasibility problem

for finding a pairwise IB scheme of a given depth t , which gives us a way to algorithmically find the smallest pairwise IB scheme for a specific (fixed) CDC. We present such a formulation in Proposition 7 in Appendix D. Additionally, Cornaz and Fonlupt [16] present a MIP formulation (with an exponential number of constraints that can be efficiently separated) to find the minimum level biclique cover of a graph.

Furthermore, we can restate the MIP formulation from [57] (which is a special case of (13) with $k=2$) in terms of biclique covers of G_S^c .

PROPOSITION 3 (Theorem 5, [57]; Theorem 1, [55]). *If $\text{CDC}(S)$ is pairwise independent branching-representable and $\{(A^j, B^j)\}_{j=1}^t$ is a biclique cover for G_S^c , then the following is an ideal formulation for $\text{CDC}(S)$:*

$$\sum_{v \in A^j} \lambda_v \leq z_j \quad \forall j \in \llbracket t \rrbracket \quad (18a)$$

$$\sum_{v \in B^j} \lambda_v \leq 1 - z_j \quad \forall j \in \llbracket t \rrbracket \quad (18b)$$

$$(\lambda, z) \in \Delta^J \times \{0, 1\}^t. \quad (18c)$$

We end the section by noting that the relation between biclique covers and independent sets has also been exploited in the study of boolean functions, particularly in the equivalence between posiforms and maximum weighted stable sets (e.g. [17, Theorem 13.16]). In fact, formulation (18) is reminiscent of formulation (13.45–13.50) in [17, Theorem 13.13]. The main difference between these formulations is that in the context of [17] the λ variables will be binary variables not constrained to lie in the unit simplex. For this reason inequalities (18a–18b) appear disaggregated in [17, Theorem 13.13] in the form $\lambda_v \leq z_j$ for all $v \in A^j$, $j \in \llbracket t \rrbracket$. However, the resulting formulation is not ideal (See [57, Section 5] for more details). Still, the combinatorial aspects of this connection could prove useful for constructing small IB schemes.

In the next section, we will explore instances where we can, in closed form, construct small (asymptotically optimal) IB schemes for families of particularly structured CDCs.

8. Illustrative examples With a framework to construct pairwise independent branching schemes for arbitrary pairwise IB-representable CDCs, we now return to some of our motivating examples. We will apply our methodology to these specific structures, and produce small, closed-form IB schemes. In particular, this allows us to construct novel, small MIP formulations for these constraints.

8.1. A simple IB scheme and its limitations To start, we show that *any* pairwise IB-representable CDC admits an IB scheme of depth $|J|$. If $|J|$ is smaller than $|S|$, this already offers a drop in size from (11). This IB scheme covers all edges incident to node with the simple biclique corresponding to the star centered at that node.

PROPOSITION 4 (Covering with Stars). *For pairwise IB-representable $\text{CDC}(S)$, a biclique cover for G_S^c is given by:*

$$A^v = \{v\}, \quad B^v = \{u \in J : \{u, v\} \in \bar{E}\} \quad \forall v \in J.$$

Proof By construction of the sets, we see that each $\{r, s\} \in \bar{E}^v \equiv A^v * B^v$ corresponds to an infeasible edge: that is, $\bar{E}^v \subseteq \bar{E}$ for each v , and so $\bigcup_{v \in J} \bar{E}^v \subseteq \bar{E}$. Furthermore, each infeasible edge $\{r, s\} \in \bar{E}$ is infeasible for levels r and s , and so $\bar{E} \subseteq \bigcup_{v \in J} \bar{E}^v$. Therefore, this construction forms a valid biclique cover of the conflict graph. \square

This gives us an upper bound of $|J|$ on the minimum depth for any pairwise IB-representable CDC. However, if we exploit the specific structure of a CDC, we can typically get much smaller formulations. For instance, consider the following two instances of the $\text{SOS3}(N)$ constraint for small values of N . First, consider the instance with $N = 6$, where $|J| = 6$ and $\mathcal{S} = \{\{1, 2, 3\}, \{2, 3, 4\}, \{3, 4, 5\}, \{4, 5, 6\}\}$. Therefore, $|\mathcal{S}| = 4$, yielding a lower bound of depth $\log_2(4) = 2$ from Proposition 1. However, there does not exist a biclique cover of depth 2 (which can be verified via Proposition 7), though one of depth 3 does exist:

$$\begin{aligned} A^1 &= \{1\}, & B^1 &= \{4, 5, 6\} \\ A^2 &= \{1, 2\}, & B^2 &= \{5, 6\} \\ A^3 &= \{1, 2, 3\}, & B^3 &= \{6\}. \end{aligned}$$

We can see the proposed IB scheme on the left side of Figure 3. For clarity, the associated MIP formulation for the CDC from Proposition 3 is

$$\begin{aligned} \lambda_1 &\leq z_1 & \lambda_4 + \lambda_5 + \lambda_6 &\leq 1 - z_1 \\ \lambda_1 + \lambda_2 &\leq z_2 & \lambda_5 + \lambda_6 &\leq 1 - z_2 \\ \lambda_1 + \lambda_2 + \lambda_3 &\leq z_3 & \lambda_6 &\leq 1 - z_3 \\ (\lambda, z) &\in \Delta^6 \times \{0, 1\}^3. \end{aligned}$$

Next, we consider $N = 10$, where we also cannot attain the $\log_2(8) = 3$ lower bound. However, a biclique for this the conflict graph of this constraint is

$$A^1 = \{1, 8, 9, 10\}, \quad B^1 = \{4, 5\} \tag{20a}$$

$$A^2 = \{1, 2, 10\}, \quad B^2 = \{5, 6, 7\} \tag{20b}$$

$$A^3 = \{1, 2, 3, 9, 10\}, \quad B^3 = \{6\} \tag{20c}$$

$$A^4 = \{1, 2, 3, 4\}, \quad B^4 = \{7, 8, 9, 10\}, \tag{20d}$$

as seen on the right side of Figure 3. The corresponding MIP formulation is

$$\begin{aligned} \lambda_1 + \lambda_8 + \lambda_9 + \lambda_{10} &\leq z_1 & \lambda_4 + \lambda_5 &\leq 1 - z_1 \\ \lambda_1 + \lambda_2 + \lambda_{10} &\leq z_2 & \lambda_5 + \lambda_6 + \lambda_7 &\leq 1 - z_2 \\ \lambda_1 + \lambda_2 + \lambda_3 + \lambda_9 + \lambda_{10} &\leq z_3 & \lambda_6 &\leq 1 - z_3 \\ \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 &\leq z_4 & \lambda_7 + \lambda_8 + \lambda_9 + \lambda_{10} &\leq 1 - z_4 \\ (\lambda, z) &\in \Delta^{10} \times \{0, 1\}^4. \end{aligned}$$

The ad-hoc construction for $\text{SOS3}(6)$ suggests a more general construction for $\text{SOS}k(N)$ when $k \leq N/2$ (assume for convenience that N is even). Consider the sets given by

$$A^j = \{1, \dots, j\} \cup \{j + N/2 + k, \dots, N\}, \quad B^j = \{j + k, \dots, j + N/2\}$$

for each $j \in \llbracket N/2 \rrbracket$. It is straightforward to see that this yields a biclique cover of the conflict graph for $\text{SOS}k(N)$ of depth $N/2$. Therefore, with this simple operation, we have constructed an ideal formulation for $\text{SOS}k(N)$ with size strictly smaller than N , the size of the naïve nonextended formulation (11).

Based on the second example (20), we know that this construction is, in general, not the smallest possible. In Section 8.4.2, we will see how we can systematically construct small biclique covers (and MIP formulations) for $\text{SOS}k(N)$ with arbitrary k and N , using techniques we will now develop.

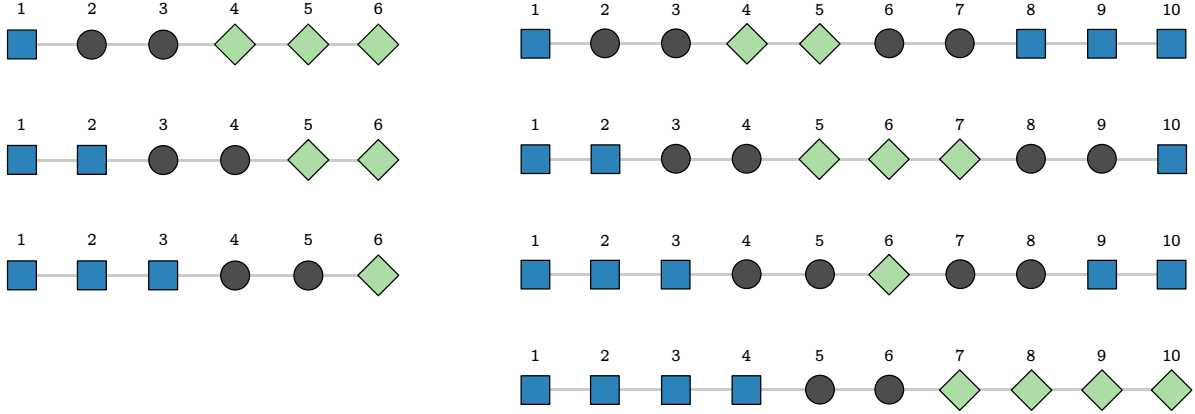


FIGURE 3. Visualizations of the biclique covers presented in the text for SOS3(6) (Left) and SOS3(10) (Right). Each row corresponds to some level j , and the elements of A^j and B^j are the blue squares and green diamonds, respectively.

8.2. Systematic construction of biclique covers As discussed in Section 6.3, there exists an IB scheme for the SOS2 constraint of optimal depth that can be constructed using a Gray code. The following proposition shows how the validity of this scheme can easily be proven by reinterpreting it via a biclique cover.

PROPOSITION 5. *Take a Gray code $\{h^i\}_{i=1}^{N-1} \subseteq \{0, 1\}^{\lceil \log_2(N-1) \rceil}$ and let $h^0 \stackrel{\text{def}}{=} h^1$ and $h^N \stackrel{\text{def}}{=} h^{N-1}$. If G_S^c is the conflict graph of SOS2(N), then a biclique cover for G_S^c of depth $\lceil \log_2(N-1) \rceil$ is given by*

$$A^j = \{\tau \in \llbracket N \rrbracket : h_j^{\tau-1} = h_j^\tau = 0\}, \quad B^j = \{\tau \in \llbracket N \rrbracket : h_j^{\tau-1} = h_j^\tau = 1\} \quad \forall j \in \llbracket \lceil \log_2(N-1) \rceil \rrbracket. \quad (21)$$

Proof For the SOS2(N) constraint we have that $\bar{E}_S = \{\{r, s\} \in \llbracket N \rrbracket^2 : r+2 \leq s\}$. Take any infeasible pair $\{r, s\} \in \bar{E}_S$. As $r+2 \leq s$, we conclude that $r-1 < r < s-1 < s$, and so it must be that $h^{r-1}, h^r \neq h^{s-1}, h^s$. The set of components which flip values between the two pairs of adjacent codes (h^{r-1}, h^r) and (h^{s-1}, h^s) is $I = \{j \in \llbracket \lceil \log_2(N-1) \rceil \rrbracket : h_j^{r-1} \neq h_j^r \text{ or } h_j^{s-1} \neq h_j^s\}$, and $|I| \leq 2$ as we have selected a Gray code. Now it must be the case that there is some component $j \in \llbracket \lceil \log_2(N-1) \rceil \rrbracket \setminus I$ wherein $h_j^{r-1} = h_j^r \neq h_j^{s-1} = h_j^s$, else we conclude that two of the vectors $h^i = h^\ell$ coincide for some $i \in \{r-1, r\}$ and $\ell \in \{s-1, s\}$, a contradiction of their uniqueness. Then $\{r, s\} \in E^j$, i.e. it is covered by the j -th level of the biclique. Furthermore, we observe that no edges of the form $\{r, r+1\}$ will be contained in the biclique cover, as it is not possible that $h_j^{r-1} = h_j^r = 0$ (resp. $= 1$) and $h_j^r = h_j^{r+1} = 1$ (resp. $= 0$) simultaneously. \square

Interestingly, we can also view this construction recursively if we use a specific Gray code known as the reflected Gray code [50]. For SOS2(2^k), we will take E^k as the edge set for the corresponding conflict graph. With $k=1$, $d=2^k=2$, $N=2^k+1=3$, and $E^1 = \{\{1, 3\}\}$. A complete biclique cover is given by the single biclique $A^{1,1} = \{1\}$ and $B^{1,1} = \{3\}$. As we see in Figure 4, we can construct a biclique cover for SOS2 with $k=2$ by stitching together two copies of the biclique. We construct two copies of the node set for $k=1$, invert the second, and identify the last node from the first set with the first node with the second set. Then we can readily construct a mapping of the biclique $(A^{1,1}, B^{1,1})$ for $k=1$ to a biclique for $k=2$, using the node identification, as $A^{2,1} = \{1, 5\}$ and $B^{2,1} = \{3\}$. This will cover all edges in E^2 with both incident nodes in the first half of the nodes, or both in the second half of the nodes (along with some other edges in E^2 , as well). To cover all edges with one adjacent node in the first half, and the other in the second half, we construct a second biclique of the form $A^{2,2} = \{1, 2\}$ and $B^{2,2} = \{4, 5\}$.

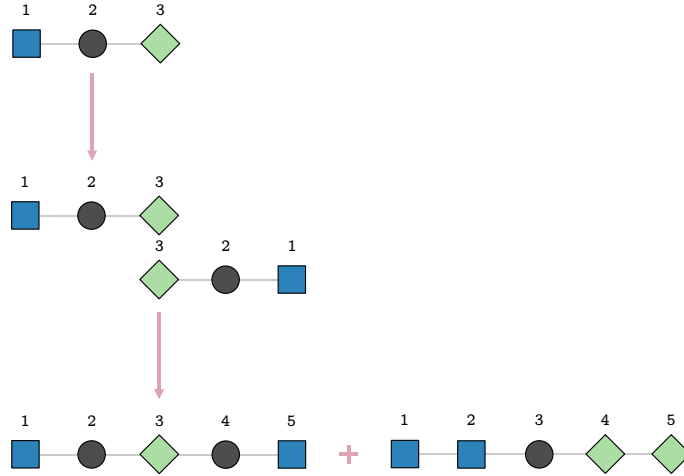


FIGURE 4. The recursive construction for biclique covers for SOS2. The first row is a single biclique that covers the conflict graph for SOS2(3) ($A^{1,1}$ in blue, $B^{1,1}$ in green). The second row shows the construction which duplicates the ground set $\{1, 2, 3\}$ and inverts the ordering on the second copy. The third row shows the identification of the nodes that yields a valid biclique for SOS2(5). This biclique is then combined with a second that covers all edges between nodes identified with the first copy and those identified with the second, giving a biclique cover for SOS2(5) with two levels.

We can repeat this construction with $k = 3$ to get the three level biclique cover

$$\begin{aligned} A^{3,1} &= \{1, 5, 9\}, & B^{3,1} &= \{3, 7\} \\ A^{3,2} &= \{1, 2, 8, 9\}, & B^{3,2} &= \{4, 5, 6\} \\ A^{3,3} &= \{1, 2, 3, 4\}, & B^{3,3} &= \{6, 7, 8, 9\}. \end{aligned}$$

We can repeat this recursively to get a biclique cover for E^{k+1} as $\{(A^{k+1,i}, B^{k+1,i})\}_{i=1}^{k+1}$, where

$$\begin{aligned} A^{k+1,i} &= \bigcup_{u \in A^{k,i}} \{u, 2^{k+1} + 2 - u\}, & B^{k+1,i} &= \bigcup_{v \in B^{k,i}} \{v, 2^{k+1} + 2 - v\} \quad \forall i \in \llbracket k \rrbracket \\ A^{k+1,k+1} &= \{1, \dots, 2^k\}, & B^{k+1,k+1} &= \{2^k + 2, \dots, 2^{k+1} + 1\}. \end{aligned}$$

In fact, we can readily state this recursive construction in a more general form, where we adapt a biclique cover for one graph into a biclique cover for another graph that is created in some specific way.

LEMMA 2. *Take some graph $G = (\llbracket m + 1 \rrbracket, E)$, and define $G^2 = (\llbracket 2m + 1 \rrbracket, E^2)$ for*

$$E^2 = E \cup \{\{2m + 2 - u, 2m + 2 - v\} : \{u, v\} \in E\} \cup (\llbracket m \rrbracket * \llbracket m + 2, 2m + 1 \rrbracket)$$

where $\llbracket a, b \rrbracket \stackrel{\text{def}}{=} \{a, \dots, b\}$. If $\{(A^j, B^j)\}_{j=1}^t$ is a biclique cover of G , then $\{(\tilde{A}^j, \tilde{B}^j)\}_{j=1}^{t+1}$ is a biclique cover of G^2 , where

$$\begin{aligned} \tilde{A}^j &= \bigcup_{u \in A^j} \{u, 2m + 2 - u\}, & \tilde{B}^j &= \bigcup_{v \in B^j} \{v, 2m + 2 - v\} \quad \forall j \in \llbracket t \rrbracket \\ \tilde{A}^{t+1} &= \{1, \dots, m\}, & \tilde{B}^{t+1} &= \{m + 2, \dots, 2m + 1\}. \end{aligned}$$

In the remainder of this work, we will see how we may apply similar graphical results to systematically construct small biclique covers for the conflict graphs of constraints by exploiting their specific structure.

8.3. Biclique covers for graph products and discretizations of multilinear terms

Consider the discretization of multilinear terms described in Section 3.4, given by $J = \prod_{i=1}^{\eta} \llbracket d_i \rrbracket$ and $\mathcal{S} = \{\prod_{i=1}^{\eta} \{k_i, k_i + 1\} : k \in \prod_{i=1}^{\eta} \llbracket d_i - 1 \rrbracket\}$. We can interpret this constraint as a η -dimensional version of the SOS2 constraint, or as the Cartesian product of η SOS2 constraints. This can be formalized through the following definition and straightforward lemma.

DEFINITION 5. The (*disjunctive*) *graph product* of a family of graphs $\{G^i = (J^i, E^i)\}_{i=1}^{\eta}$ is $\bigvee_{i=1}^{\eta} G^i \stackrel{\text{def}}{=} (J_P, E_P)$, where $J_P = \prod_{i=1}^{\eta} J^i$ and

$$E_P = \{\{u, v\} \in J_P * J_P : \exists i \in \llbracket \eta \rrbracket \text{ s.t. } \{u_i, v_i\} \in E^i\}.$$

LEMMA 3. Let $J = \prod_{i=1}^{\eta} \llbracket d_i \rrbracket$ and $\mathcal{S} = \{\prod_{i=1}^{\eta} \{k_i, k_i + 1\} : k \in \prod_{i=1}^{\eta} \llbracket d_i - 1 \rrbracket\}$ be the η -dimensional discretization of multilinear terms, and $G_{\mathcal{S}}^c$ be the corresponding conflict graph. If G^i is the conflict graph of SOS2(d_i) for each $i \in \llbracket \eta \rrbracket$, then $G_{\mathcal{S}}^c = \bigvee_{i=1}^{\eta} G^i$.

Using this characterization, we can easily construct an IB scheme for discretizations of multilinear terms by taking the graph products of IB schemes for the SOS2 constraint.

LEMMA 4. Take a family of graphs $\{G^i = (J^i, E^i)\}_{i=1}^{\eta}$, and a biclique cover $\{\{\tilde{A}^{i,j}, \tilde{B}^{i,j}\}_{j=1}^{t_i}\}$ for each G^i . Then a biclique cover for $\bigvee_{i=1}^{\eta} G^i$ is given by $\bigcup_{i=1}^{\eta} \{(A^{i,j}, B^{i,j})\}_{j=1}^{t_i}$, where

$$A^{i,j} = \left(\prod_{\ell=1}^{i-1} J^{\ell} \right) \times \tilde{A}^{i,j} \times \left(\prod_{\ell=i+1}^{\eta} J^{\ell} \right), \quad B^{i,j} = \left(\prod_{\ell=1}^{i-1} J^{\ell} \right) \times \tilde{B}^{i,j} \times \left(\prod_{\ell=i+1}^{\eta} J^{\ell} \right) \quad \forall i \in \llbracket \eta \rrbracket, j \in \llbracket t_i \rrbracket.$$

COROLLARY 3. Let $J = \prod_{i=1}^{\eta} \llbracket d_i \rrbracket$ and $\mathcal{S} = \{\prod_{i=1}^{\eta} \{k_i, k_i + 1\} : k \in \prod_{i=1}^{\eta} \llbracket d_i - 1 \rrbracket\}$ describe the η -dimensional discretization of multilinear terms, and take $G_{\mathcal{S}}^c$ as its conflict graph. If for each $i \in \llbracket \eta \rrbracket$ we have a biclique cover $\{\{\tilde{A}^{i,j}, \tilde{B}^{i,j}\}_{j=1}^{t_i}\}$ for the conflict graph of SOS2(d_i), then a biclique cover for $G_{\mathcal{S}}^c$ of depth $\sum_{i=1}^{\eta} t_i$ is given by $\bigcup_{i=1}^{\eta} \{(A^{i,j}, B^{i,j})\}_{j=1}^{t_i}$, where

$$A^{i,j} = \{x \in J : x_i \in \tilde{A}^{i,j}\}, \quad B^{i,j} = \{x \in J : x_i \in \tilde{B}^{i,j}\} \quad \forall i \in \llbracket \eta \rrbracket, j \in \llbracket t_i \rrbracket.$$

In particular, if we take $\{h^{i,j}\}_{j=1}^{d_i-1} \subseteq \{0, 1\}^{\lceil \log_2(d_i-1) \rceil}$ as a Gray code for each $i \in \llbracket \eta \rrbracket$, where $h^{i,0} \stackrel{\text{def}}{=} h^{i,1}$ and $h^{i,d_i} \stackrel{\text{def}}{=} h^{i,d_i-1}$, then a biclique cover for $G_{\mathcal{S}}^c$ of depth $\sum_{i=1}^{\eta} \lceil \log_2(d_i - 1) \rceil$ is given by:

$$\hat{A}^{i,j} = \{x \in J : \exists \gamma \text{ s.t. } x_i = \gamma, h_j^{i,\gamma-1} = h_j^{i,\gamma} = 0\}, \quad \hat{B}^{i,j} = \{x \in J : \exists \gamma \text{ s.t. } x_i = \gamma, h_j^{i,\gamma-1} = h_j^{i,\gamma} = 1\}$$

for each $i \in \llbracket \eta \rrbracket$ and $j \in \llbracket \lceil \log_2(d_i - 1) \rceil \rrbracket$.

We note that, since $|\mathcal{S}| = \prod_{i=1}^{\eta} (d_i - 1)$, by Proposition 1 this construction yields a formulation that is asymptotically optimal (with respect to number of auxiliary binary variables) for any possible MIP formulation, up to an additive factor of at most η .

Furthermore, we can specialize this to the bilinear case studied by Misener et al. [41].

COROLLARY 4. There exists a biclique cover for a the grid discretization of a bilinear function ($\eta = 2$) with $d_2 = 1$ of depth $\lceil \log_2(d_1 - 1) \rceil$.

This result yields an ideal MIP formulation for the outer-approximation of bilinear terms with $\lceil \log_2(m) \rceil$ auxiliary binary variables, $2(m + 1)$ auxiliary continuous variables (the λ variables, one for element in J), and $2\lceil \log_2(m) \rceil$ general inequality constraints. In contrast, the logarithmic formulation from Misener et al. [41] has $\lceil \log_2(m) \rceil$ auxiliary binary variables, $2\lceil \log_2(m) \rceil + 1$ auxiliary continuous variables, at least $2\lceil \log_2(m) \rceil + 6$ general inequality constraints, and is not ideal in general (see Appendix A). Therefore, we gain an ideal formulation with a naturally induced constraint branching at the price of a modest number of additional auxiliary continuous variables. Furthermore, our formulation generalizes readily to discretization along the second dimension ($d_2 > 1$), for non-uniform discretizations, and for higher dimensional multilinear functions ($\eta > 2$).

8.4. Completing biclique covers via graph unions Another useful graphical technique for our heuristic constructions will be to combine together biclique covers, each of which is designed to cover a substructure of the constraint. For example, the conflict graph of a grid triangulation of the plane is equivalent to the conflict graph of the 2-dimensional grid discretization of multilinear terms, with one extra edge added for each subrectangle in the grid. Therefore, a biclique cover of a grid triangulation can be obtained from a biclique cover of 2-dimensional discretization of multilinear terms (i.e. from Corollary 3) by completing it with some number additional bicliques that cover those extra edges. This construction can be formalized in the following way.

DEFINITION 6. The *graph union* of a family of graphs $\{G^i = (J^i, E^i)\}_{i=1}^{\eta}$ is $\bigcup_{i=1}^{\eta} G^i \stackrel{\text{def}}{=} (J_U, E_U)$, where $J_U = \bigcup_{i=1}^{\eta} J^i$ and $E_U = \bigcup_{i=1}^{\eta} E^i$.

LEMMA 5. Take a family of graphs $\{G^i = (J^i, E^i)\}_{i=1}^{\eta}$ and a corresponding biclique cover $\{(A_i^j, B_i^j)\}_{j=1}^{t_i}$ of G^i for each $i \in \llbracket \eta \rrbracket$. Then $\bigcup_{i=1}^{\eta} \{(A_i^j, B_i^j)\}_{j=1}^{t_i}$ is a biclique cover of $\bigcup_{i=1}^{\eta} G^i$.

We can apply Lemma 5 to construct biclique covers for the grid triangulations depicted in Figure 1. First, we apply the biclique cover construction from Corollary 3 to cover all edges not sharing a subrectangle. This is depicted in the first two subfigures of each row in Figure 5. To cover the remaining 4 edges created by the triangulation, we see that the number of additional levels needed is dependent on the combinatorial structure. Additionally, in all three cases we can verify through Proposition 7 that the resulting biclique cover is of the smallest possible depth. The first

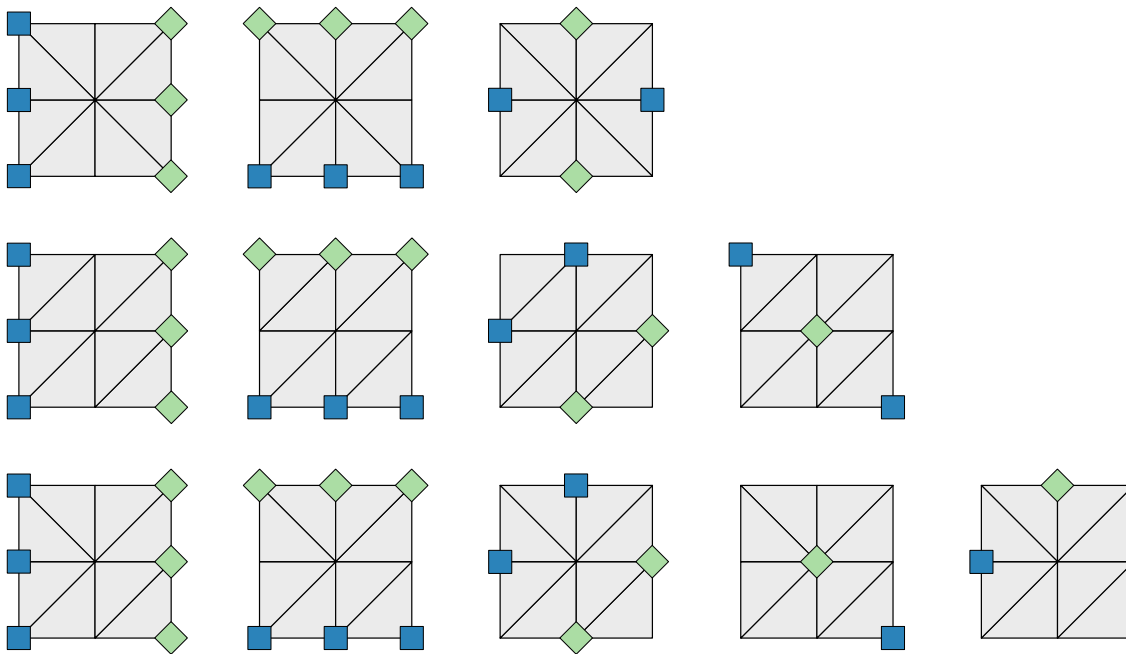


FIGURE 5. Independent branching schemes for the three triangulations presented in Figure 1, each given its own row. The sets A^j and B^j are given by the blue boxes and green diamonds vertices, respectively, in the i -th subfigure of the corresponding row.

example is the “Union Jack” triangulation [51] for $N = M = 3$ and the results in [57] show that for this triangulation the biclique cover from Corollary 3 can be completed with a single additional biclique cover for any N and M . The second triangulation is a K1 triangulation [34] for $N = M = 3$, and an early version of [53] showed that for this triangulation the biclique can always be completed with two additional bicliques (See Section 8.5 for a generalization of these results).

In contrast, for generic triangulations such as the third one, it was not previously known if the biclique cover can always be completed with fewer than the trivial $(M - 1)(N - 1)$ levels needed to cover each “diagonal” edge with its own additional biclique. First, we can adapt Proposition 4 to cover the extra edges with stars, but in general this will result in $\Theta(M \cdot N)$ stars, and hence the same number of additional levels. To reduce this, we need a way to *stick* the stars together into more complicated bicliques. In section 8.5 we will see how we may use graph colorings for a broad class of triangulations (subsuming the Union Jack and K1 triangulations as special cases), to cover the extra edges with either one or two additional bicliques. In general, it turns out that we may cover the remaining edges for any grid triangulation with a constant number of additional levels by applying the simple following lemma.

LEMMA 6. *Let $\{(A^j, B^j)\}_{j=1}^t$ be a family of bicliques of a graph G . If (A^k, B^ℓ) is also a biclique of G for each $k, \ell \in [t]$, then $(\bigcup_{j=1}^t A^j, \bigcup_{j=1}^t B^j)$ is a biclique of G .*

The strength of Lemma 6 comes from the fact that many CDCs of practical interest have a local structure (i.e. sets in \mathcal{S} have small cardinality, or, equivalently, the minimum degree of the conflict graph is close to the total number of nodes). In this case, the condition of Lemma 6 will hold for families of stars centered at nodes that are located “sufficiently far apart.”

8.4.1. Grid triangulations of the plane We may now present a biclique cover construction for generic grid triangulations, with no further assumptions on the structure of the triangles such as in [55, 57], whose depth scales like $\log_2(M) + \log_2(N) + \mathcal{O}(1)$. In the same way as depicted in Figure 5, we construct the biclique cover by using Lemma 5 to complete the construction of Corollary 3. For this, we will use the following corollary of Lemma 6 that shows how to combine certain stars centered at sufficiently separated nodes.

COROLLARY 5. *Take a regular grid $J = [M] \times [N]$, let \mathcal{S} be a grid triangulation of $[1, M] \times [1, N]$, and take $G_{\mathcal{S}}^c = (J, \bar{E})$ as its conflict graph. For all $w \in J$, define $A(w) \stackrel{\text{def}}{=} \{w\}$ and $B(w) \stackrel{\text{def}}{=} \{w + v : v \in \{-1, 1\}^2, \{w, w + v\} \in \bar{E}\}$. Then*

$$\left(\bigcup_{w \in J \cap (u + 3\mathbb{Z}^2)} A(w), \bigcup_{w \in J \cap (u + 3\mathbb{Z}^2)} B(w) \right)$$

is a biclique of $G_{\mathcal{S}}^c$ for any $u \in J$.

Proof Direct from Lemma 6 by taking $u \in J$ and the family of bicliques $\{(A(w), B(w))\}_{w \in J \cap (u + 3\mathbb{Z}^2)}$ and noting that, if $u, v \in J \cap (u + 3\mathbb{Z}^2)$, then $\|u - v\|_{\infty} \geq 3$, and so $(A(u), B(v))$ is also a biclique for $G_{\mathcal{S}}^c$. \square

Figure 6 shows two possible bicliques that can be obtained from Corollary 5.

We can now use Lemma 5 and Corollary 5, along with the biclique cover derived in Corollary 3, to obtain a biclique cover for any triangulation with an asymptotically optimal number of levels.

THEOREM 4. *Take $J = [M] \times [N]$, and let \mathcal{S} be a grid triangulation of $[1, M] \times [1, N]$ with $J = \bigcup \{S \in \mathcal{S}\}$. Take $G_{\mathcal{S}}^c = (J, \bar{E})$ as its conflict graph. Presume that $\{(\tilde{A}^{1,j}, \tilde{B}^{1,j})\}_{j=1}^{t_1}$ and $\{(\tilde{A}^{2,j}, \tilde{B}^{2,j})\}_{j=1}^{t_2}$ are biclique covers for the conflict graphs of the $SOS2(M)$ and $SOS2(N)$ constraints, respectively. Furthermore, define*

$$\begin{aligned} A^{3,u} &= J \cap (u + 3\mathbb{Z}^2) \\ B^{3,u} &= \bigcup_{w \in J \cap (u + 3\mathbb{Z}^2)} \left\{ w + v : v \in \{-1, 1\}^2, \{w, w + v\} \in \bar{E} \right\} \end{aligned}$$

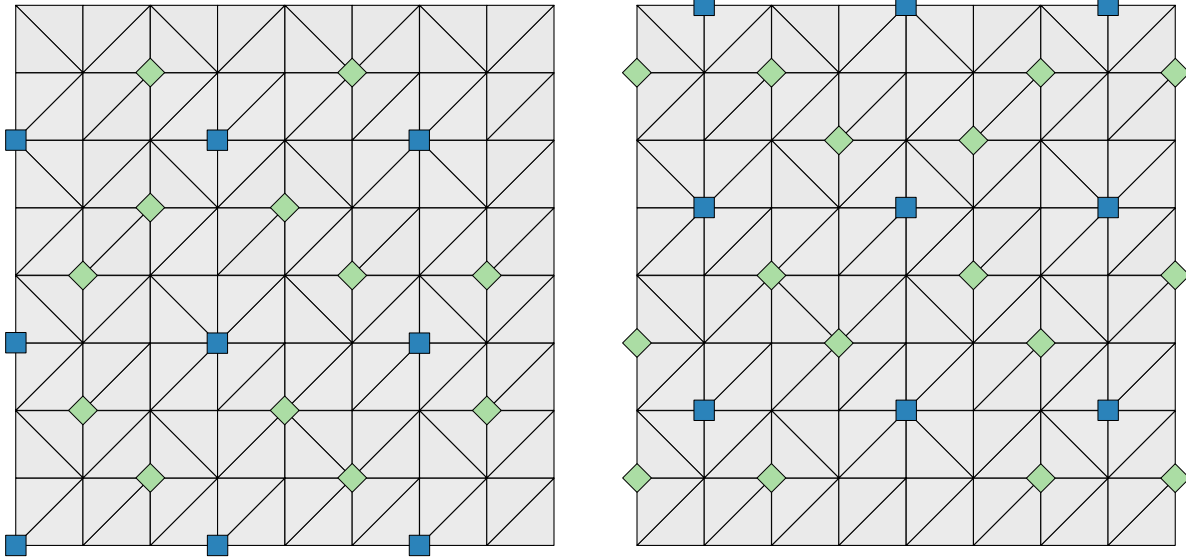


FIGURE 6. Two bicliques constructed via Corollary 5 for a grid triangulation with $M = N = 8$. On the left the construction follows by taking $u = (1, 1)$; on the right, with $u = (2, 3)$. For each level, the sets $A(u)$ and $B(u)$ are given by the blue squares and green diamonds, respectively.

for each $u \in \{0, 1, 2\}^2$. Then $\{(A^{1,j}, B^{1,j})\}_{j=1}^{t_1} \cup \{(A^{2,j}, B^{2,j})\}_{j=1}^{t_2} \cup \{(A^{3,u}, B^{3,u})\}_{u \in \{0,1,2\}^2}$ is a biclique cover for G_S^c , where

$$\begin{aligned} A^{1,j} &= \tilde{A}^{1,j} \times \llbracket N \rrbracket, & B^{1,j} &= \tilde{B}^{1,j} \times \llbracket N \rrbracket, \\ A^{2,j'} &= \llbracket M \rrbracket \times \tilde{A}^{2,j'}, & B^{2,j'} &= \llbracket M \rrbracket \times \tilde{B}^{2,j'}, \end{aligned}$$

for each $j \in \llbracket t_1 \rrbracket$ and $j' \in \llbracket t_2 \rrbracket$.

In particular, if $\{h^{1,i}\}_{i=1}^{M-1} \subseteq \{0, 1\}^{\lceil \log_2(M-1) \rceil}$ and $\{h^{2,i}\}_{i=1}^N \subseteq \{0, 1\}^{\lceil \log_2(N-1) \rceil}$ are Gray codes, where $h^{1,0} \stackrel{\text{def}}{=} h^{1,1}$, $h^{1,M} \stackrel{\text{def}}{=} h^{1,M-1}$, $h^{2,0} \stackrel{\text{def}}{=} h^{2,1}$, and $h^{2,N} \stackrel{\text{def}}{=} h^{2,N-1}$, then a biclique cover of G_S^c of depth $\lceil \log_2(M-1) \rceil + \lceil \log_2(N-1) \rceil + 9$ is given by:

$$A^{1,\ell} = \{(x, y) \in J : h_\ell^{1,x-1} = h_\ell^{1,x} = 1\} \quad (22a)$$

$$B^{1,\ell} = \{(x, y) \in J : h_\ell^{1,x-1} = h_\ell^{1,x} = 0\} \quad (22b)$$

$$A^{2,\ell'} = \{(x, y) \in J : h_\ell^{2,y-1} = h_\ell^{2,y} = 1\} \quad (22c)$$

$$B^{2,\ell'} = \{(x, y) \in J : h_\ell^{2,y-1} = h_\ell^{2,y} = 0\} \quad (22d)$$

$$A^{3,u} = J \cap (u + 3\mathbb{Z}^2) \quad (22e)$$

$$B^{3,u} = \bigcup_{w \in J \cap (u + 3\mathbb{Z}^2)} \{w + v : v \in \{-1, 1\}^2, \{w, w + v\} \in \bar{E}\} \quad (22f)$$

for all $\ell \in \llbracket \lceil \log_2(M-1) \rceil \rrbracket$, $\ell' \in \llbracket \lceil \log_2(N-1) \rceil \rrbracket$, and $u \in \{0, 1, 2\}^2$.

Proof Let $G^x \stackrel{\text{def}}{=} (\llbracket M \rrbracket, E^x)$ and $G^y \stackrel{\text{def}}{=} (\llbracket N \rrbracket, E^y)$ be the conflict graphs for $\text{SOS2}(M)$ and $\text{SOS2}(N)$, respectively. Furthermore, let

$$G^3 \stackrel{\text{def}}{=} \bigcup_{u \in \{0,1,2\}^2} (J, A^{3,u} * B^{3,u}) = \left(J, \bigcup_{u \in \{0,1,2\}^2} (A^{3,u} * B^{3,u}) \right).$$

Then we see that $G_S = (G^x \times G^y) \cup G^3$ by noting that all *diagonal* edges of E (i.e. those of the form $\{w, w + v\} \in E$ for $w \in J$ and $v \in \{-1, 1\}^2$) are included in G^3 , and observing that G^3 is a subgraph of G_S^c . The result then follows from Lemma 4, Lemma 5, and Corollary 5. \square

By referring to Proposition 1, we recover a $\lceil \log_2(2(M-1)(N-1)) \rceil \geq \lceil \log_2(M-1) \rceil + \lceil \log_2(N-1) \rceil$ lower bound on the depth of any biclique cover for a grid triangulation, and see that our construction yields a MIP formulation that is within a constant additive factor of the smallest possible. Finally, we note that similarly to the results in [55, 57] for the Union Jack triangulation, formulation (18) for biclique cover (22) can provide a significant computational advantage for general grid triangulations [26].

8.4.2. SOS k In this subsection, we will see how we may use the graph union construction to produce an IB scheme for SOS $k(N)$ of depth $\log_2(N/k) + \mathcal{O}(k)$, for any k and N . Similar to the construction for grid triangulations, we first construct an initial family of bicliques based on the SOS2 constraint. Next, we expand this onto a larger node set by the graph product construction. Finally, we complete the biclique cover by combining a family of sufficiently separated stars. For grid triangulations, this approach meant applying SOS2 constraints horizontally and vertically, and taking a graph product of the two. One way to interpret this is as an SOS2 constraint applied to groups of aggregated nodes in the ground set (e.g. when SOS2 is applied horizontally, we group all elements with the same horizontal coordinate into a single group). For the SOS $k(N)$ constraint, we will apply the SOS2 constraint to the groups obtained by partitioning the N original ground elements into $\lceil N/k \rceil$ subsets of k consecutive elements. The following simple lemma shows how this grouping can also be represented through a graph product. For the remainder of the section, we assume that N/k is integer; if this is not true, we artificially introduce $\lceil N/k \rceil k - N$ nodes such that this is the case, construct the formulation in Theorem 5, and remove the artificial nodes from the formulation afterwards.

LEMMA 7. *Let $J = \llbracket N \rrbracket$, $k \leq N$, \mathcal{S} correspond to the SOS $k(N)$ constraint, and $G_s^c = (J, \bar{E})$ be the corresponding conflict graph. Let $G^1 = (\llbracket N/k \rrbracket, E^1)$ be the conflict graph for SOS2(N/k) and $G^2 = (\{0, \dots, k-1\}, \emptyset)$ be the empty graph on k nodes. Then $G^1 \times G^2$ is isomorphic to a subgraph $\hat{G} = (J, \hat{E})$ of G_s^c wherein $\hat{E} \subseteq \bar{E}$, and each edge $\{u, v\} \in \bar{E}$ with $|u - v| \geq 2k$ is contained $\{u, v\} \in \hat{E}$.*

Proof Let $G^1 \times G^2 = (J', E')$. Consider the bijection $f : J \rightarrow J'$ given by $f(u) = (\text{div}(u, k), \text{mod}(u, k))$, where $\text{div}(u, k) = \lfloor u/k \rfloor$ and $\text{mod}(u, k) = u - k \text{div}(u, k)$ are the quotient and remainder of the division of u by k , so that $f^{-1}(m, r) = km + r$. We have that $\{(m, r), (m', r')\} \in E'$ if and only if $\{m, m'\} \in E^1$, which in turn is equivalent to $|m - m'| \geq 2$. Therefore, for any $\{(m, r), (m', r')\} \in E'$, we have

$$|f^{-1}(m, r) - f^{-1}(m', r')| = |(km + r) - (km' + r')| = |k(m - m') + (r - r')| \geq k|m - m'| + |r - r'| \geq 2k,$$

and hence $\{f^{-1}(m, r), f^{-1}(m', r')\} \in \bar{E}$, i.e. $\hat{E} \subseteq \bar{E}$. For the second condition, see that if $u, v \in J$ are such that $|u - v| \geq 2k$, then $|\text{div}(u) - \text{div}(v)| \geq 2$, and therefore $\{f(u), f(v)\} \in E'$. \square

We can then cover the remaining edges with the following bicliques obtained by stitching together families of sufficiently separated stars.

COROLLARY 6. *Let $J = \llbracket N \rrbracket$, $k \leq N$, \mathcal{S} correspond to the SOS $k(N)$ constraint, and G_s^c be the corresponding conflict graph. For all $w \in J$, define $A(w) \stackrel{\text{def}}{=} \{w\}$ and $B(w) \stackrel{\text{def}}{=} \{u \in J : k \leq |u - w| < 2k\}$. Then*

$$\left(\bigcup_{w \in J \cap (u + 3k\mathbb{Z})} A(w), \bigcup_{w \in J \cap (u + 3k\mathbb{Z})} B(w) \right)$$

is a biclique of G_s^c for any $u \in J$.

Proof Direct from Lemma 6 by considering the family of bicliques $\{(A(w), B(w))\}_{w \in J \cap (u + 3k\mathbb{Z})}$ and noting that, for distinct $u, v \in J \cap (u + 3k\mathbb{Z})$, $|u - v| \geq 3k$, and so $(A(u), B(v))$ is also a biclique for G_s^c . \square

Finally, we can combine both classes of bicliques with Lemma 5 to construct a complete biclique cover for $\text{SOS}k(N)$. See Figure 7 for an example of the resulting construction.

THEOREM 5. *Let $J = \llbracket N \rrbracket$, $k \leq N$, \mathcal{S} correspond to the $\text{SOS}k(N)$ constraint on J , and $G_{\mathcal{S}}^c$ be the corresponding conflict graph. Let $\{(\tilde{A}^{1,j}, \tilde{B}^{1,j})\}_{j=1}^{t_1}$ be a biclique cover for the conflict graph of the $\text{SOS}2(N/k)$ constraint, and take*

$$A^{2,j'} = \bigcup_{i=0}^{\lceil \frac{N}{3k} \rceil} \{\tau \in J : \tau = j' + (3i - 3)k\}$$

$$B^{2,j'} = \bigcup_{i=0}^{\lceil \frac{N}{3k} \rceil} \{\tau \in J : j' + (3i - 2)k \leq \tau \leq j' + (3i - 1)k\}$$

for all $j' \in \llbracket 3k \rrbracket$. Then $\{(A^{1,j}, B^{1,j})\}_{j=1}^t \cup \{(A^{2,j'}, B^{2,j'})\}_{j'=1}^{\llbracket 3k \rrbracket}$ is a biclique cover for $G_{\mathcal{S}}^c$, where

$$A^{1,j} = \left\{ \tau \in J : \lceil \tau/k \rceil \in \tilde{A}^{1,j} \right\}, \quad B^{1,j} = \left\{ \tau \in J : \lceil \tau/k \rceil \in \tilde{B}^{1,j} \right\},$$

for each $j \in \llbracket t \rrbracket$.

In particular, if $\{h^i\}_{i=1}^{\lceil N/k \rceil - 1} \subseteq \{0, 1\}^{\lceil \log_2(\lceil N/k \rceil - 1) \rceil}$ is a Gray code where $h^0 \stackrel{\text{def}}{=} h^1$ and $h^{\lceil N/k \rceil} \stackrel{\text{def}}{=} h^{\lceil N/k \rceil - 1}$, then a biclique cover of $G_{\mathcal{S}}^c$ of depth $\lceil \log_2(\lceil N/k \rceil - 1) \rceil + 3k$ is given by $\{(A^{1,j}, B^{1,j})\}_{j=1}^{\lceil \log_2(\lceil N/k \rceil - 1) \rceil} \cup \{(A^{2,j'}, B^{2,j'})\}_{j'=1}^{\llbracket 3k \rrbracket}$, where

$$A^{1,j} = \left\{ \tau \in J : h_j^{\lceil \tau/k \rceil - 1} = h_j^{\lceil \tau/k \rceil} = 0 \right\}, \quad B^{1,j} = \left\{ \tau \in J : h_j^{\lceil \tau/k \rceil - 1} = h_j^{\lceil \tau/k \rceil} = 1 \right\}$$

for all $j \in \llbracket \lceil \log_2(\lceil N/k \rceil - 1) \rrbracket \rrbracket$.

Proof Take $G^1 \stackrel{\text{def}}{=} (\llbracket N/k \rrbracket, E^1)$ as the conflict graph for $\text{SOS}2(N/k)$, $G^2 \stackrel{\text{def}}{=} (\{0, \dots, k-1\}, \emptyset)$ as the empty graph on k nodes, and $G^3 \stackrel{\text{def}}{=} \left(J, \bigcup_{j'=1}^{3k} A^{2,j'} * B^{2,j'} \right)$. Let \hat{G} be the subgraph of $G_{\mathcal{S}}^c$ from Lemma 7, which is isomorphic to $G^1 \times G^2$ through the bijection $g : \llbracket N/k \rrbracket \times \{0, \dots, k-1\} \rightarrow J$ with $g(m, r) = km + r$. Then we have that $G_{\mathcal{S}}^c = \hat{G} \cup G^3$, after applying Lemma 7 and using the fact that the edges of G^3 contain the edges of $G_{\mathcal{S}}^c$ not included in \hat{G} . The result then follows from Lemma 4, Lemma 5, Lemma 7, and Corollary 5. \square

We note that, when $k = \mathcal{O}(\log(N))$, this biclique cover yields a MIP formulation that is asymptotically tight (with respect to the number of auxiliary binary variables) with our lower bound of $\lceil \log_2(N - k + 1) \rceil$ from Proposition 1. We can also show an absolute lower bound of depth k for any biclique cover for $\text{SOS}k$. This implies that when $k = \omega(\log(N))$, although the formulation from Theorem 5 is not tight with respect to the lower bound from Proposition 1, it is asymptotically the smallest possible formulation in the pairwise IB framework.

PROPOSITION 6. *Any biclique cover for the conflict graph of $\text{SOS}k(N)$ must have depth at least $\min\{k, N - k\}$.*

Proof Define $\gamma \stackrel{\text{def}}{=} \min\{k, N - k\}$ and consider any possible biclique cover $\{(A^j, B^j)\}_{j=1}^t$. The biclique cover must separate the edges $\{(\tau, \tau + k)\}_{\tau=1}^{\gamma}$. Consider a level j of the biclique cover that contains edge $\{\tau, \tau + k\}$ for some $\tau \in \llbracket \gamma \rrbracket$; w.l.o.g., $\tau \in A^j$ and $\tau + k \in B^j$. Consider the possibility that the same level j separates another such edge in the set, e.g. $(\tau', \tau' + k)$ for $\tau' \in \llbracket \gamma \rrbracket$, where w.l.o.g. $\tau < \tau'$. That would imply that either $\tau' \in A^j$ or $\tau' \in B^j$. In the case that $\tau' \in A^j$, we have that \bar{E}^j contains the edge $\{\tau', \tau + k\}$. However, since $|(\tau + k) - \tau'| = \tau + k - \tau' < \tau + k - \tau = k$, this implies that the biclique cover separates a feasible edge, a contradiction. In the case where $\tau' \in B^j$, we have that \bar{E}^j contains the edge $\{\tau, \tau'\}$, and as $\tau' - \tau < k \leq \gamma$ from the definition of our set of edges, a similar argument holds. Therefore, each edge $\{(\tau, \tau + k)\}_{\tau=1}^{\gamma}$ must be uniquely contained in some level of the biclique cover, giving the result. \square

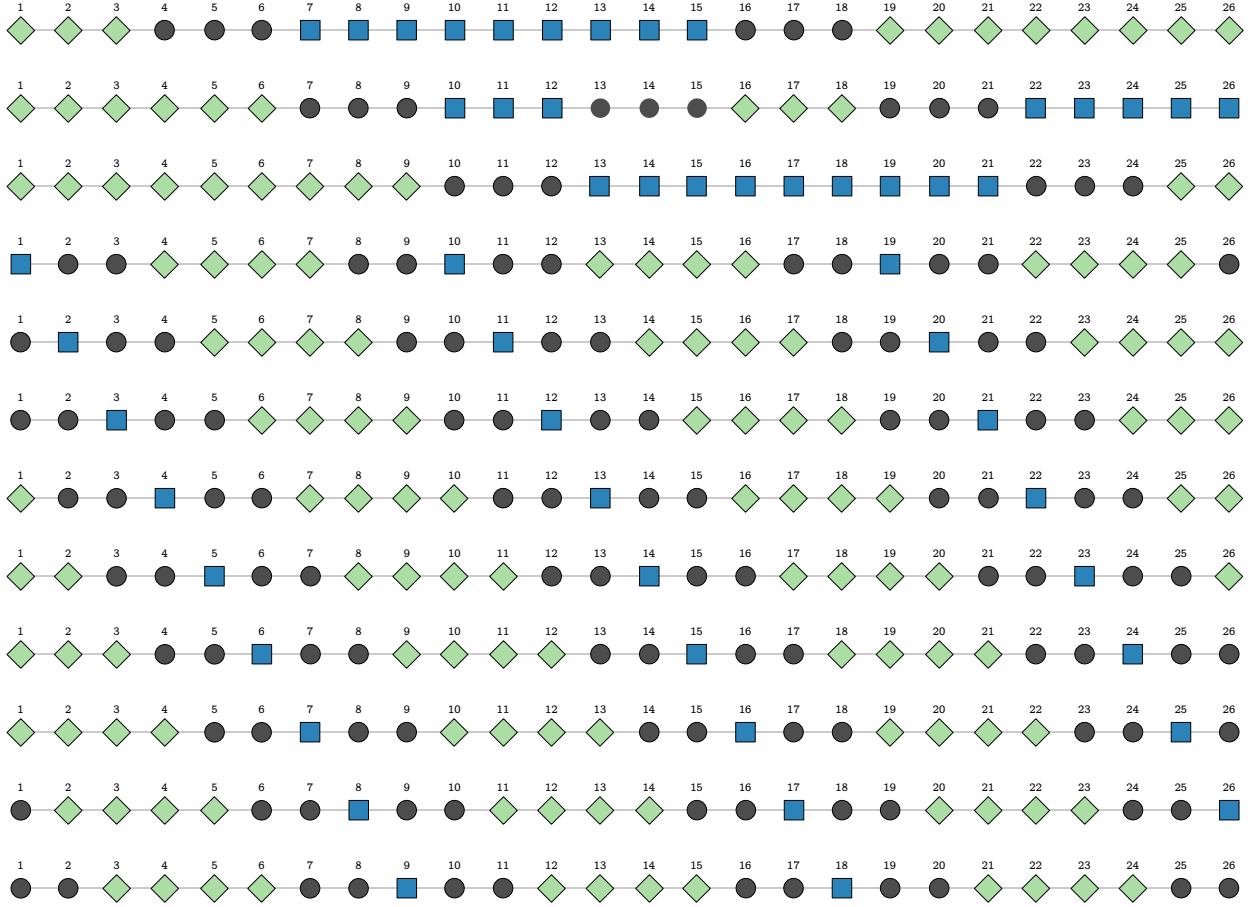


FIGURE 7. Visualizations of the biclique cover from the proof of Theorem 5 for SOS3(26). Each row corresponds to some level j , and the sets A^j and B^j are the blue squares and green diamonds, respectively. The first three rows correspond to the “first stage” of the biclique cover $\{(A^{1,j}, B^{1,j})\}_{j=1}^3$, and the second nine correspond to the “second stage” $\{(A^{2,j}, B^{2,j})\}_{j=1}^9$.

Furthermore, when $k = \lfloor N/2 \rfloor$, this proposition gives a lower bound on the depth of a biclique cover that is asymptotically tight with the upper bound of N from Proposition 4. In other words, in this particular regime, we have that the SOS k constraint admits a pairwise IB-based formulation, but only one that is relatively large ($\Omega(|S|) = \Omega(|J|)$ auxiliary binary variables and constraints).

8.5. Covering edges with a chromatic characterization of bicliques Our final contribution is to adapt a result of Cornaz and Fonlupt [16] that gives a chromatic characterization of the set of edges that can be covered by a biclique.

THEOREM 6 ([16]). *Take the graph $G = (J, \bar{E})$, along with some edge subset $\bar{F} \subseteq \bar{E}$. Define $V(\bar{F}) = \bigcup \{\{u, v\} \in \bar{F}\}$ as all nodes incident to \bar{F} , and take $F = [(\bar{F} * \bar{F}) \setminus \bar{E}]$ as all pairs incident to the edges $V(\bar{F})$ not contained in \bar{E} . Define both $E' = F \cup \bar{F}$ and $p: E' \rightarrow \{0, 1\}$ such that $p(e) = \mathbb{1}[e \in \bar{E}]$ ⁸. Finally, take $\mathcal{C}(E')$ as the family of all cycles in G' . Then the following are equivalent:*

1. *There exists a biclique (A, B) of G covering $(V(\bar{F}), \bar{F})$.*
2. *For all $C \in \mathcal{C}(E')$, $\sum_{u \in C} p(u)$ is even.*
3. *There exists some $f: V(\bar{F}) \rightarrow \{0, 1\}$ such that*

⁸ Where $\mathbb{1}[e \in \bar{E}] = 1$ if $e \in \bar{E}$ and $\mathbb{1}[e \in \bar{E}] = 0$ otherwise.

- $f(u) = f(v)$ for all $\{u, v\} \in F$,
- $f(u) \neq f(v)$ for all $\{u, v\} \in \bar{F}$, and
- $(\{u \in V(\bar{F}) : f(u) = 0\}, \{u \in V(\bar{F}) : f(u) = 1\})$ is a biclique of G covering $(V(\bar{F}), \bar{F})$.

Using this result, we can sometimes exploit the structure of a grid triangulations to obtain biclique covers that are smaller than those derived in Theorem 4.

THEOREM 7. *Take a regular grid $J = \llbracket M \rrbracket \times \llbracket N \rrbracket$, the sets \mathcal{S} given by a grid triangulation of $[1, M] \times [1, N]$ with $\bigcup \{S \in \mathcal{S}\} = J$, and take $G_{\mathcal{S}}^c = (J, \bar{E})$ to be the corresponding conflict graph. Furthermore, let $\{(\tilde{A}^{1,j}, \tilde{B}^{1,j})\}_{j=1}^{t_1}$ and $\{(\tilde{A}^{2,j}, \tilde{B}^{2,j})\}_{j=1}^{t_2}$ be biclique covers for the conflict graphs of the SOS2(M) and SOS2(N) constraints, respectively. Also define $J_{\text{even}} \stackrel{\text{def}}{=} \{u \in J : u_1 \equiv u_2 \pmod{2}\}$ and $J_{\text{odd}} \stackrel{\text{def}}{=} \{u \in J : u_1 \not\equiv u_2 \pmod{2}\}$ as the set of nodes whose sum of components is either even and odd, respectively. For each $s \in \{\text{even}, \text{odd}\}$, let $E_s \stackrel{\text{def}}{=} \{\{u, v\} \in [J_s]^2 : \|u - v\|_{\infty} = 1\}$, $\bar{F}_s \stackrel{\text{def}}{=} E_s \cap \bar{E}$, and $F_s \stackrel{\text{def}}{=} ([V(\bar{F}_s)]^2 \cap E_s) \setminus \bar{E}$.*

If, for each $s \in \{\text{even}, \text{odd}\}$, there exists $f_s : J_s \rightarrow \{0, 1\}$ such that

$$f_s(u) = f_s(v) \quad \forall \{u, v\} \in F_s \quad \text{and} \quad f_s(u) \neq f_s(v) \quad \forall \{u, v\} \in \bar{F}_s, \quad (23)$$

then $\{(A^{1,j}, B^{1,j})\}_{j=1}^{t_1} \cup \{(A^{2,j}, B^{2,j})\}_{j=1}^{t_2} \cup \{(A^{3,i}, B^{3,i})\}_{i \in [2]}$ is a biclique cover for $G_{\mathcal{S}}^c$, where

$$\begin{aligned} A^{1,j} &= \tilde{A}^{1,j} \times \llbracket N \rrbracket, & B^{1,j} &= \tilde{B}^{1,j} \times \llbracket N \rrbracket, \\ A^{2,j'} &= \llbracket M \rrbracket \times \tilde{A}^{2,j'}, & B^{2,j'} &= \llbracket M \rrbracket \times \tilde{B}^{2,j'}, \\ A^{3,s} &= \{u \in J_s : f_s(u) = 0\}, & B^{3,s} &= \{u \in J_s : f_s(u) = 1\} \end{aligned}$$

for each $j \in [t_1]$, $j' \in [t_2]$, and $s \in \{\text{even}, \text{odd}\}$.

If we select $\{(\tilde{A}^{1,j}, \tilde{B}^{1,j})\}_{j=1}^{t_1}$ and $\{(\tilde{A}^{2,j}, \tilde{B}^{2,j})\}_{j=1}^{t_2}$ to correspond to the Gray code construction for SOS2 in Proposition 5, then the resulting biclique cover has depth $\lceil \log_2(M-1) \rceil + \lceil \log_2(N-1) \rceil + 2$. Furthermore, if for some $s \in \{\text{even}, \text{odd}\}$ we have that $f_s(u) = \alpha$ for all $u \in J_s$ and some constant α , we may reduce the depth to $\lceil \log_2(M-1) \rceil + \lceil \log_2(N-1) \rceil + 1$.

Finally, if we fix $s \in \{\text{even}, \text{odd}\}$ and $r \in \{\text{even}, \text{odd}\} \setminus \{s\}$, a sufficient condition for the existence of $f_s : J_s \rightarrow \{0, 1\}$ satisfying (23) is that

$$d_{F_r}(u) \stackrel{\text{def}}{=} |\{e \in F_r : \exists v \in J_r \text{ s.t. } e = \{u, v\}\}| \text{ is even} \quad \forall u \in J_r \cap \llbracket 2, M-1 \rrbracket \times \llbracket 2, N-1 \rrbracket. \quad (24)$$

Proof Let $G^x \stackrel{\text{def}}{=} (\llbracket M \rrbracket, E^x)$ and $G^y \stackrel{\text{def}}{=} (\llbracket N \rrbracket, E^y)$ be the conflict graphs for the SOS2(M) and SOS2(N) constraints, respectively. Furthermore, let $G^3 \stackrel{\text{def}}{=} (\llbracket M \rrbracket \times \llbracket N \rrbracket, (A^{3,\text{even}} * B^{3,\text{even}}) \cup (A^{3,\text{odd}} * B^{3,\text{odd}}))$. Then we may see that $G_{\mathcal{S}}^c = (G^x \times G^y) \cup G^3$ by noting that all *diagonal* edges of \bar{E} (i.e. those of the form $\{w, w+v\} \in \bar{E}$ for $w \in J$ and $v \in \{-1, 1\}^2$) are included in G^3 , and that G^3 is a subgraph of $G_{\mathcal{S}}^c$. The first part of the theorem then follows from Lemma 4, Lemma 5, and Theorem 6.

For the sufficient condition, w.l.o.g. consider the case where $s = \text{even}$ and $r = \text{odd}$. Define $p_{\text{even}} : E_{\text{even}} \rightarrow \{0, 1\}$ as $p_{\text{even}}(e) = \mathbf{1}[e \in \bar{E}]$. The result will follow from Theorem 6 by showing that (24) satisfies condition 2 in the equivalence of the theorem. Let $E'_{\text{even}} = F_{\text{even}} \cup \bar{F}_{\text{even}}$, and assume for contradiction that there exists $C \in \mathcal{C}(E'_{\text{even}})$ such that $\sum_{u \in C} p_{\text{even}}(u)$ is odd. If $|C| = 4$, we may assume without loss of generality that $V(C) = \{u, u + (1, 1), u + (1, -1), u + (2, 0)\} \subset J_{\text{even}}$ for some $u \in J_{\text{even}}$. Then $v = u + (0, 1) \in J_{\text{odd}}$ is such that $d_{E_{\text{odd}}}(v)$ is odd, a contradiction of (24). If $|C| > 4$, note that $C \in \mathcal{C}(E_{\text{even}})$ and that there exists $e \in C$ such that $p_{\text{even}}(e) = 1$. In addition, there exists $C_0 \in \mathcal{C}(E_{\text{even}})$ such that $e \in C_0$, $|C_0| = 4$, $C_1 = (C_0 \cup C) \setminus (C_0 \cap C) \in \mathcal{C}(E_{\text{even}})$ and $\text{Conv}(C_1) \subsetneq \text{Conv}(C)$. If $\sum_{u \in C_0} p_{\text{even}}(u)$ is odd, we may make the same argument above as $|C_0| = 4$ to derive a contradiction of (24). If not, then C_1 and C have the same parity, and therefore $\sum_{u \in C_1} p_{\text{even}}(u)$ is odd. We may then repeat this shrinking procedure recursively on C_1 until either the 4-cycle C_0 has odd parity, or C_1 is itself a 4-cycle. In either case, we have a 4-cycle with odd parity, which implies some $u \in J_{\text{odd}}$ that violates (24), giving the result. \square

We note that we may use this coloring characterization to recover the biclique covers for both the Union Jack and K1 triangulation example in Figure 5. For the Union Jack example, we have that

$$\bar{F}_{\text{even}} = \emptyset, \quad \bar{F}_{\text{odd}} = \{\{(1, 2), (2, 1)\}, \{(2, 1), (3, 2)\}, \{(1, 2), (2, 3)\}, \{(2, 3), (3, 2)\}\},$$

and so we may apply the simplification in Theorem 7 to construct a biclique cover of depth $\log_2(2) + \log_2(2) + 1 = 3$. Indeed, the original formulation of Vielma and Nemhauser [57] for the Union Jack triangulation can be reinterpreted analogously through the chromatic characterization of Theorem 7.

For the K1 triangulation example, we have

$$\bar{F}_{\text{even}} = \{\{(1, 3), (2, 2)\}, \{(2, 2), (3, 1)\}\}, \quad \bar{F}_{\text{odd}} = \{\{(1, 2), (2, 1)\}, \{(2, 3), (3, 2)\}\},$$

giving a biclique cover of depth $\log_2(2) + \log_2(2) + 2 = 4$.

Furthermore, we close by noting that the sufficient condition in Theorem 7 is, in general, not necessary. For example, in Figure 8 we see a grid triangulation that does not satisfy (24), but for which there exists a coloring given by $\{f_{\text{even}}, f_{\text{odd}}\}$ that satisfies (23). That is, $(2, 3)$ has odd degree (i.e. $d_{F_{\text{odd}}}((2, 3)) = 1$), and so the sufficient condition is not satisfied. However, a coloring satisfying the conditions of Theorem 7 exists. This offers a generalization of the result that originally appeared in a preliminary version of [53], which only showed the sufficient condition (24).

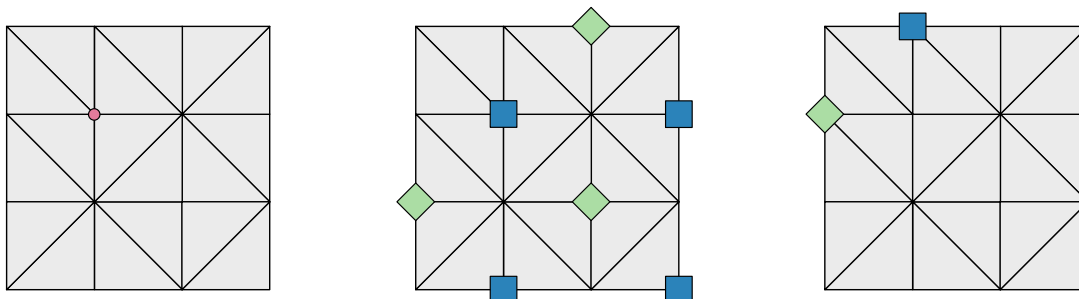


FIGURE 8. (Left) A grid triangulation that does not satisfy the sufficient condition of Theorem 7, as $(2, 3)$ (red circle) has odd degree. However, there is a coloring of the form described in Theorem 7, leading to two bicliques (Center and Right) that cover all “diagonal” edges of the conflict graph.

Acknowledgements This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1122374, and by the National Science Foundation under Grant No. CMMI-1351619. The authors would like to thank Yves Crama for pointing out the relation between biclique covers, independent sets, and boolean functions.

References

- [1] Achterberg, Tobias, Timo Berthold, Thorsten Koch, Kati Wolter. 2008. *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems: 5th International Conference, CPAIOR 2008 Paris, France, May 20-23, 2008 Proceedings*, chap. Constraint integer programming: A new approach to integrate CP and MIP. Springer Berlin Heidelberg, 6–20.
- [2] Achterberg, Tobias, Thorsten Koch, Alexander Martin. 2005. Branching rules revisited. *Operations Research Letters* **33**(1) 42–54.

- [3] Adams, Warren P., Stephen M. Henry. 2012. Base-2 expansions for linearizing products of functions of discrete variables. *Operations Research* **60**(6) 1477–1490.
- [4] Aichholzer, Oswin, Franz Aurenhammer, Ferran Hurtado, Hannes Krasser. 2003. Towards compatible triangulations. *Theoretical Computer Science* **296**(1) 3–13.
- [5] Androulakis, Ioannis, Costas D. Maranas. 1995. α BB: A global optimization method for general constrained nonconvex problems. *Journal of Global Optimization* **7**(4) 337–363.
- [6] Applegate, J. A., R. K. Wood. 2000. Explicit-constraint branching for solving mixed-integer programs. M. Laguna, J. L. González, eds., *Computing tools for modeling, optimization, and simulation: interfaces in computer science and operations research*, *Operations research/computer science interfaces series*, vol. 12. Kluwer, 245–261.
- [7] Apt, Krzysztof R. 2003. *Principles of Constraint Programming*. Cambridge University Press.
- [8] Balas, Egon. 1985. Disjunctive programming and a hierarchy of relaxations for discrete optimization problems. *SIAM J. Alg. Disc. Meth.* **6**(3) 466–486.
- [9] Beale, E. M. L., J. A. Tomlin. 1970. Special facilities in a general mathematical programming system for non-convex problems using ordered sets of variables. J. Lawrence, ed., *OR 69: Proceedings of the Fifth International Conference on Operational Research*. Tavistock Publications, 447–454.
- [10] Bellingham, John Saunders. 2002. Coordination and control of uav fleets using mixed-integer linear programming. Master’s thesis, Massachusetts Institute of Technology.
- [11] Bertsimas, Dimitris, Romy Shioda. 2009. Algorithm for cardinality-constrained quadratic optimization. *Comput. Optim. Appl.* **43** 1–22.
- [12] Bienstock, Daniel. 1996. Computational study of a family of mixed-integer quadratic programming problems. *Mathematical Programming* **74** 121–140.
- [13] Bixby, R., E. Rothberg. 2007. Progress in computational mixed integer programming—A look back from the other side of the tipping point. *Annals of Operations Research* **149** 37–41.
- [14] Chang, T.-J., N. Meade, J.E. Beasley, Y.M. Sharaiha. 2000. Heuristics for cardinality constrained portfolio optimization. *Computers and Operations Research* **27** 1271–1302.
- [15] Conforti, Michele, Gérard Cornuéjols, Giacomo Zambelli. 2014. *Integer Programming*. Springer.
- [16] Cornaz, Denis, Jean Fonlupt. 2006. Chromatic characterization of biclique covers. *Discrete Mathematics* **306** 495–507.
- [17] Crama, Yves, Peter L Hammer. 2011. *Boolean functions: Theory, algorithms, and applications*. Cambridge University Press.
- [18] de Farias Jr., I. R., E. L. Johnson, G. L. Nemhauser. 2001. Branch-and-cut for combinatorial optimization problems without auxiliary binary variables. *The Knowledge Engineering Review* **16** 25–39.
- [19] Deits, Robin, Russ Tedrake. 2015. Efficient mixed-integer planning for UAVs in cluttered environments. *IEEE International Conference on Robotics and Automation*. 42–49.
- [20] Fishburn, Peter C., Peter L. Hammer. 1996. Bipartite dimensions and bipartite degrees of graphs. *Discrete Mathematics* **160** 127–148.
- [21] Floudas, Christodoulos A., Xiaoxia Lin. 2005. Mixed integer linear programming in process scheduling: Modeling, algorithms, and applications. *Annals of Operations Research* **139** 131–162.
- [22] Foulds, L.R., D. Haugland, K. Jörnsten. 1992. A bilinear approach to the pooling problem. *Optimization* **24**(1-2) 165–180.
- [23] Garey, Michael R., David S. Johnson. 1979. *Computers and Intractability*. W. H. Freeman and Company.
- [24] Gruber, Hermann, Markus Holzer. 2007. *Inapproximability of nondeterministic state and transition complexity assuming $P \neq NP$* , *Lecture Notes in Computer Science*, vol. 4588. Springer US, 205–216.
- [25] Hooker, John N. 2002. Logic, optimization, and constraint programming. *INFORMS Journal on Computing* **14**(4) 295–321.
- [26] Huchette, J., J. P. Vielma. 2017. Mixed-integer formulations for piecewise linear functions: Modern approaches. 2017 Mixed Integer Programming Workshop, June 19–22, 2017, Montréal, Québec, Canada, Poster. URL <http://www.mit.edu/~huchette/files/mip-2017-poster.pdf>.

- [27] Ismael R. de Farias, Jr., George L. Nemhauser. 2003. A polyhedral study of the cardinality constrained knapsack problem. *Math. Program., Ser. A* **96** 439–467.
- [28] Jaffar, Joxan, Michael J. Maher. 1994. Constraint logic programming: A survey. *J. Logic Programming* **20** 503–581.
- [29] Jeroslow, R.G., J.K. Lowe. 1984. Modelling with integer variables. *Mathematical Programming Study* **22** 167–184.
- [30] Jünger, Michael, Thomas Liebling, Denis Naddef, George Nemhauser, William Pulleyblank, Gerhard Reinelt, Giovanni Rinaldi, Laurence Wolsey. 2010. *50 years of integer programming 1958-2008*. Springer.
- [31] Keha, A. B., I. R. de Farias, G. L. Nemhauser. 2004. Models for representing piecewise linear cost functions. *Operations Research Letters* **32** 44–48.
- [32] Keha, A. B., I. R. de Farias, G. L. Nemhauser. 2006. A branch-and-cut algorithm without binary variables for nonconvex piecewise linear optimization. *Operations Research* **54** 847–858.
- [33] Kondili, E., C. C. Pantelides, R. W. H. Sargent. 1993. A general algorithm for short-term scheduling of batch operations—I. MILP formulation. *Computers and Chemical Engineering* **17**(2) 211–227.
- [34] Kuhn, Harold W. 1960. Some combinatorial lemmas in topology. *IBM Journal of Research and Development* **4**(5) 518–528.
- [35] Land, Ailsa, Alison Doig. 1960. An automatic method for solving discrete programming problems. *Econometrica* **28**(3) 497–520.
- [36] Luedtke, James, Mahdi Namazifar, Jeff Linderoth. 2012. Some results on the strength of relaxations of multilinear functions. *Mathematical Programming* **136**(2) 325–351.
- [37] Martin, A., M. Moller, S. Moritz. 2006. Mixed integer models for the stationary case of gas network optimization. *Mathematical Programming* **105** 563–582.
- [38] McCormick, Garth. 1976. Computability of global solutions to factorable nonconvex programs: Part I—Convex underestimating problems. *Mathematical Programming* **10** 147–175.
- [39] Mellinger, Daniel, Alex Kushleyev, Vijay Kumar. 2012. Mixed-integer quadratic program trajectory generation for heterogeneous quadrotor teams. *IEEE International Conference on Robotics and Automation*. 477–483.
- [40] Misener, Ruth, Christodoulos Floudas. 2012. Global optimization of mixed-integer quadratically-constrained quadratic programs (MIQCQP) through piecewise-linear and edge-concave relaxations. *Mathematical Programming, Series B* **136** 155–182.
- [41] Misener, Ruth, Jeffrey P. Thompson, Christodoulos A. Floudas. 2011. APOGEE: Global optimization of standard, generalized, and extended pooling problems via linear and logarithmic partitioning schemes. *Computers and Chemical Engineering* **35** 876–892.
- [42] Muldoon, Frank M., Warren P. Adams, Hanif D. Sherali. 2013. Ideal representations of lexicographic orderings and base-2 expansions of integer variables. *Operations Research Letters* **41** 32–39.
- [43] Orlin, James. 1977. Contentment in graph theory: Covering graphs with cliques. *Indagationes Mathematicae (Proceedings)* **80**(5) 406–424.
- [44] Ostrowski, James, Jeff Linderoth, Fabrizio Rossi, Stefano Smriglio. 2011. Orbital branching. *Math. Program., Ser. A* **126** 147–178.
- [45] Prodan, Ionela, Florin Stoican, Sorin Olaru, Silviu-Iulian Niculescu. 2012. Enhancements on the hyperplanes arrangements in mixed-integer programming techniques. *J. Optim. Theory Appl.* **154** 549–572.
- [46] Prodan, Ionela, Florin Stoican, Sorin Olaru, Silviu-Iulian Niculescu. 2016. *Mixed-Integer Representations in Control Design: Mathematical Foundations and Applications*. SpringerBriefs in Electrical and Computer Engineering, Springer.
- [47] Quesada, Ignacio, Ignacio E. Grossmann. 1995. A global optimization algorithm for linear fractional and bilinear programs. *Journal of Global Optimization* **6**(1) 39–76.
- [48] Ryan, D. M., B. A. Foster. 1981. An integer programming approach to scheduling. *Computer Scheduling of Public Transport* 269–280.

- [49] Sahinidis, Nikolaos V. 1996. BARON: A general purpose global optimization software package. *Journal of Global Optimization* **8** 201–205.
- [50] Savage, Carla. 1997. A survey of combinatorial Gray codes. *SIAM Review* **39**(4) 605–629.
- [51] Todd, Michael J. 1977. Union Jack triangulations. *Fixed Points: Algorithms and Applications* 315–336.
- [52] Vielma, Juan Pablo. 2015. Mixed integer linear programming formulation techniques. *SIAM Review* **57**(1) 3–57.
- [53] Vielma, Juan Pablo. 2017. Embedding formulations and complexity for unions of polyhedra. *To appear in Management Science*.
- [54] Vielma, Juan Pablo, Shabbir Ahmed, George Nemhauser. 2008. A lifted linear programming branch-and-bound algorithm for mixed-integer conic quadratic programs. *INFORMS Journal on Computing* **20**(3) 438–450.
- [55] Vielma, Juan Pablo, Shabbir Ahmed, George Nemhauser. 2010. Mixed-integer models for nonseparable piecewise-linear optimization: Unifying framework and extensions. *Operations Research* **58**(2) 303–315.
- [56] Vielma, Juan Pablo, Ahmet B. Keha, George L. Nemhauser. 2008. Nonconvex, lower semicontinuous piecewise linear optimization. *Discrete Optimization* **5**(2) 467–488.
- [57] Vielma, Juan Pablo, George Nemhauser. 2011. Modeling disjunctive constraints with a logarithmic number of binary variables and constraints. *Mathematical Programming, Series A* **128** 49–72.
- [58] Wicaksono, Danan S., Iftekhhar A. Karimi. 2008. Piecewise MILP under- and overestimators for global optimisation of bilinear programs. *AIChE Journal* **54**(4) 991–1008.
- [59] Yildiz, Sercan, Juan Pablo Vielma. 2013. Incremental and encoding formulations for mixed integer programming. *Operations Research Letters* **41** 654–658.
- [60] Ziegler, Gunter. 2007. *Lectures on Polytopes*. Springer.

Appendix A: Logarithmic formulation from Misener et al. [41] is not ideal We show that the logarithmic formulation (16) from Misener et al. [41] is not, in general, ideal. Using their notation, we take $N_P = 3$, $x^L = y^L = 0$, and $x^U = y^U = 3$ (and so $a = 1$). Then formulation (16) is

$$\lambda_1 + 2\lambda_2 \leq x \tag{25a}$$

$$x \leq 1 + \lambda_1 + 2\lambda_2 \tag{25b}$$

$$1 + \lambda_1 + 2\lambda_2 \leq 3 \tag{25c}$$

$$\Delta y_1 \leq 3\lambda_1 \tag{25d}$$

$$\Delta y_2 \leq 3\lambda_2 \tag{25e}$$

$$\Delta y_1 = y - s_1 \tag{25f}$$

$$\Delta y_2 = y - s_2 \tag{25g}$$

$$s_1 \leq 3(1 - \lambda_1) \tag{25h}$$

$$s_2 \leq 3(1 - \lambda_2) \tag{25i}$$

$$z \geq \Delta y_1 + 2\Delta y_2 \tag{25j}$$

$$z \geq 3x + (y - 3) + (\Delta y_1 - 3\lambda_1) + 2(\Delta y_2 - 3\lambda_2) \tag{25k}$$

$$z \leq y + \Delta y_1 + 2\Delta y_2 \tag{25l}$$

$$z \leq 3x + (\Delta y_1 - 3\lambda_1) + 2(\Delta y_2 - 3\lambda_2) \tag{25m}$$

$$\lambda \in \{0, 1\}^2 \tag{25n}$$

$$\Delta y \in [0, 3]^2 \tag{25o}$$

$$s \in [0, 3]^2 \tag{25p}$$

$$(x, y) \in [0, 3] \times [0, 3]. \tag{25q}$$

The feasible point for the relaxation $x = 3$, $y = 3$, $z = 9$, $\lambda = (1, 0.5)$, $\Delta y = (3, 1.5)$, and $s = (0, 1.5)$ is a fractional extreme point, showing that the formulation is not ideal. Indeed, it satisfies at equality the set of linear independent constraints of the relaxation given by $x \leq 3$, $y \leq 3$, $\lambda_1 \leq 1$, $\Delta y_1 \leq 3$, $s_1 \geq 0$, (25b), (25e), (25g) and (25k).

Appendix B: Proof of Proposition 1 First, we present a more general lemma.

LEMMA 8. *If there do not exist polyhedra $\{Q^i\}_{i=1}^{d'}$ with $d' < d$ and $\bigcup_{i=1}^d P^i = \bigcup_{i=1}^{d'} Q^i$, then any binary MIP formulation for $\bigcup_{i=1}^d P^i$ must have at least $\lceil \log_2(d) \rceil$ binary variables.*

Proof Presume that formulation F takes the form (8). For each $h \in \{0, 1\}^{n_3}$, consider the preimage $\text{Pre}(h) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^{n_1} : \exists y \in \mathbb{R}^{n_2} \text{ s.t. } (x, y, h) \in F\}$; it is clear from the definition of F that this set is polyhedral. Furthermore, we need that $\bigcup_{h \in \{0, 1\}^{n_3}} \text{Pre}(h) = \bigcup_{i=1}^d P^i$. If the condition holds, then we have that the cardinality of the index set of the left side (2^{n_3}) must be at least d , which implies the result. \square

We may apply this lemma in the case where our irredundancy assumption holds.

Proof of Proposition 1 For each $S \in \mathcal{S}$, take $\lambda^S = \frac{1}{|S|} \sum_{v \in S} \mathbf{e}^v$, where $\mathbf{e}^v \in \{0, 1\}^J$ is the unit vector for component v . Assume that there is some Q^i as in Lemma 8 such that $\lambda^S, \lambda^{S'} \in Q^i$ for two $S, S' \in \mathcal{S}$. By convexity, $\frac{1}{2}(\lambda^S + \lambda^{S'}) \in Q^i$ as well. But this implies that there is a point in Q^i with support over $S \cup S'$, which would violate the irredundancy of S and S' , meaning that this cannot yield a formulation for $\bigcup_{i=1}^d P^i$. Therefore, each of the d points λ^S must be contained uniquely in some Q^i , and we may apply Lemma 8 for the result. \square

Appendix C: Independence in formulation-induced branching schemes As discussed in [57, Section 3], the connection between MIP formulations and branching schemes for CDCs can be used to explain in what sense an independent branching scheme is “independent.” As noted in Section 5.1, the branching scheme on λ induced by formulation (13) is precisely the multi-variable branching associated to the corresponding independent branching scheme. In contrast, formulations that are not based on IB schemes (e.g. (10)) do not necessarily induce a multi-variable branching. However, we can interpret the induced effect on the λ variables as a multi-way branching scheme that fixes the λ variables in a *non-independent* way.

For example, consider SOS2(5) (i.e. $\mathcal{S} = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 5\}\}$). For this particular instance and for $\{h^S\}_{S \in \mathcal{S}}$ given by $h^{\{1,2\}} = (1, 1)$, $h^{\{2,3\}} = (1, 0)$, $h^{\{3,4\}} = (0, 1)$, and $h^{\{4,5\}} = (0, 0)$, formulation (10) is given by

$$\lambda_1 = \gamma_1^{\{1,2\}}, \quad \lambda_2 = \gamma_2^{\{1,2\}} + \gamma_2^{\{2,3\}}, \quad \lambda_3 = \gamma_3^{\{2,3\}} + \gamma_3^{\{3,4\}}, \quad (26a)$$

$$\lambda_4 = \gamma_4^{\{3,4\}} + \gamma_4^{\{4,5\}}, \quad \lambda_5 = \gamma_5^{\{4,5\}} \quad (26b)$$

$$\gamma_1^{\{1,2\}} + \gamma_2^{\{1,2\}} + \gamma_2^{\{2,3\}} + \gamma_3^{\{2,3\}} + \gamma_3^{\{3,4\}} + \gamma_4^{\{3,4\}} + \gamma_4^{\{4,5\}} + \gamma_5^{\{4,5\}} = 1 \quad (26c)$$

$$\gamma_1^{\{1,2\}} + \gamma_2^{\{1,2\}} + \gamma_2^{\{2,3\}} + \gamma_3^{\{2,3\}} = z_1 \quad (26d)$$

$$\gamma_1^{\{1,2\}} + \gamma_2^{\{1,2\}} + \gamma_3^{\{3,4\}} + \gamma_4^{\{3,4\}} = z_2 \quad (26e)$$

$$\gamma_v^S \geq 0 \quad \forall v, S \quad (26f)$$

$$(\lambda, z) \in \Delta^5 \times \{0, 1\}^2 \quad (26g)$$

and a pairwise IB-based formulation (simplified slightly from (13)) is

$$\lambda_1 + \lambda_2 \leq z_1, \quad \lambda_4 + \lambda_5 \leq 1 - z_1 \quad (27a)$$

$$\lambda_3 \leq z_2, \quad \lambda_1 + \lambda_5 \leq 1 - z_2 \quad (27b)$$

$$(\lambda, z) \in \Delta^5 \times \{0, 1\}^2. \quad (27c)$$

In Figure 9, we see the first two levels of the branch-and-bound trees for both formulations for the cases where we choose either z_1 or z_2 for which to branch on first. We observe that, for formulation (26), the variables λ_v that a given branching decision is able to prove are zero depends on the previous branching decisions in the branch-and-bound tree, while this is not the case for the independent branching formulation (27). For example, if we first branch down on z_2 ($z_2 \leq 0$), we are able to prove that $\lambda_1 = 0$. If we choose instead to branch down on z_1 ($z_1 \leq 0$), we are able to

prove that $\lambda_1 = \lambda_2 = 0$. However, if we branch down on z_2 and then branch down on z_1 , we prove that $\lambda_1 = \lambda_2 = 0$, but we are also able to prove that $\lambda_3 = 0$, which we could not prove without the combination of the two branching decisions. Indeed, we see that regardless of the branching decision we make, we will not be able to prove that $\lambda_3 = 0$ until the second level of the branching tree with formulation (26).

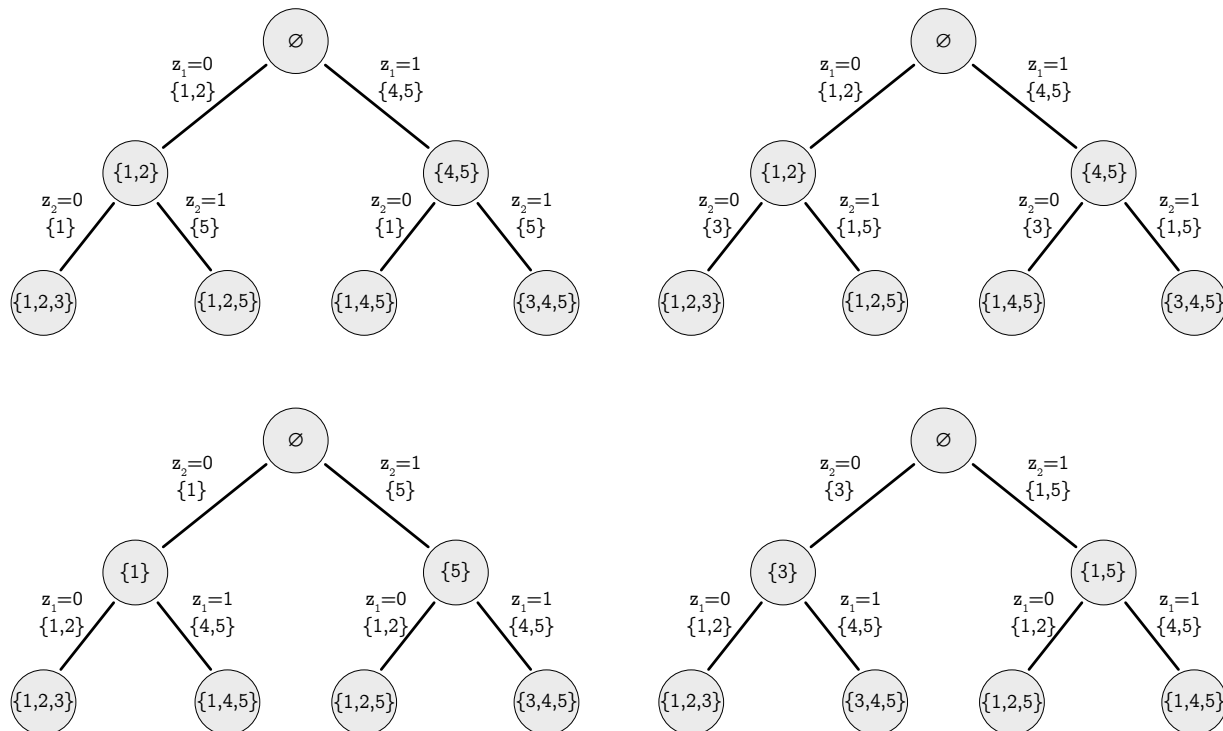


FIGURE 9. The branch-and-bound trees for (26) (Left) and (27) (Right), when z_1 is first to branch on, and then z_2 (Top row), and when z_2 is first to branch on, and then z_1 (Bottom row). Inside each node is the set $I \subset \llbracket 5 \rrbracket$ of all components v for which the algorithm has been able to prove that $\lambda_v = 0$ at this point in the algorithm via branching decisions. The text on the lines show the current branching decision (e.g. $z_2 \geq 1$), and the set of variables λ_v for which the (a) subproblem is able to prove that $\lambda_v = 0$ *independently* of any other branching decisions (e.g. $z_2 \geq 1$ is the only additional branching constraint added to the original relaxation). This figure is adapted from [57, Figure 2].

Contrastingly, each branching decision with the independent branching formulation (27) is able to fix components of λ to zero, *independent* of the location in the tree and of the previous branching decisions. For example, branching down or up on z_2 is always able to prove either $\lambda_3 = 0$ or $\lambda_1 = \lambda_5 = 0$, independently. Consequentially, for every component $v \in J$, there exists a branching decision that is able to prove that $\lambda_v = 0$ at the first level of the branch-and-bound tree, which is not the case with formulation (26) and $v = 3$, as mentioned above. Having this independence property is a restriction on the branching scheme, but has the potential to simplify branching rules (i.e. choosing which variable z_i to branch on), a notoriously difficult and computationally important part of the algorithmic performance of a MIP solver (see, for example, [2]). Furthermore, we see that independent branching rules guarantee that the solver can prove any component of λ is zero at the very beginning of the tree.

Finally, we note that MIP formulations that are not independent branching formulations can still exhibit the independent branching behavior. For example, if we had selected the encoding

$\{h^S\}_{S \in \mathcal{S}}$ to be given by given by $h^{\{1,2\}} = (1, 1)$, $h^{\{2,3\}} = (1, 0)$, $h^{\{3,4\}} = (0, 0)$, and $h^{\{4,5\}} = (0, 1)$, formulation (10) would satisfy the independent branching property. However, independent branching formulations provide an immediate proof that the property holds, which is not the case for general MIP formulations.

Appendix D: Proposition 7

PROPOSITION 7. *A biclique cover of depth t exists for the conflict graph $G_S^c = (J, \bar{E})$ of pairwise IB-representable CDC(S) if and only if the following admits a feasible solution:*

$$\left. \begin{array}{l} z_j^{r,s} \leq x_j^r + x_j^s \\ z_j^{r,s} \leq x_j^r + y_j^r \\ z_j^{r,s} \leq x_j^s + y_j^s \\ z_j^{r,s} \leq y_j^r + y_j^s \\ z_j^{r,s} \geq x_j^r + y_j^s - 1 \\ z_j^{r,s} \geq x_j^s + y_j^r - 1 \end{array} \right\} \quad \forall j \in \llbracket t \rrbracket, \forall \{r, s\} \in [J]^2 \quad (28a)$$

$$x_j^r + y_j^r \leq 1 \quad \forall j \in \llbracket t \rrbracket, \forall r \in J \quad (28b)$$

$$\sum_{j=1}^t z_j^{r,s} \geq 1 \quad \forall \{r, s\} \in \bar{E} \quad (28c)$$

$$\sum_{j=1}^t z_j^{r,s} = 0 \quad \forall \{r, s\} \in [J]^2 \setminus \bar{E} \quad (28d)$$

$$x^r \in \{0, 1\}^t \quad \forall r \in J \quad (28e)$$

$$y^r \in \{0, 1\}^t \quad \forall r \in J \quad (28f)$$

$$z^{r,s} \in \{0, 1\}^t \quad \forall \{r, s\} \in [J]^2. \quad (28g)$$

Moreover, for any feasible solution (x, y, z) , a biclique cover for G_S^c is given by $A^j = \{r \in J : x_j^r = 1\}$ and $B^j = \{r \in J : y_j^r = 1\}$ for each $j \in \llbracket t \rrbracket$.

Proof The interpretation of the decision variables is:

$$x_j^r = \mathbb{1} [r \in A^j] \quad (29a)$$

$$y_j^r = \mathbb{1} [r \in B^j] \quad (29b)$$

$$z_j^{r,s} = \mathbb{1} [x_j^r = y_j^s = 1 \text{ or } x_j^s = y_j^r = 1]. \quad (29c)$$

That is, $z_j^{r,s} = 1$ iff level i separates infeasible edge $\{r, s\} \in \bar{E}$, which is enforced via (28a-28b). To show that the existence of a biclique cover implies that (28) is feasible, you may consider the proposed solution (29) and see that it is feasible for (28).

To show that a feasible solution maps to a biclique cover, consider some (x, y, z) feasible for (28), and the corresponding sets $A^j = \{r \in J : x_j^r = 1\}$ and $B^j = \{r \in J : y_j^r = 1\}$ for each $j \in \llbracket t \rrbracket$. Inequalities (28b) ensure that $A^j \cap B^j = \emptyset$ for each $j \in \llbracket t \rrbracket$. Constraints (28d) ensure that $A^j * B^j \subseteq \bar{E}$ for each $j \in \llbracket t \rrbracket$. Therefore, each (A^j, B^j) is a biclique of G_S^c . Furthermore, (28c) ensures that there is at least one level j that separates each infeasible edge $\{r, s\} \in \bar{E}$. Therefore, $\{(A^j, B^j)\}_{j=1}^t$ is a biclique cover of G_S^c . \square