

# A geometric way to build strong mixed-integer programming formulations

Joey Huchette<sup>1</sup> and Juan Pablo Vielma<sup>2</sup>

<sup>1</sup> Rice University, [joehuchette@rice.edu](mailto:joehuchette@rice.edu)

<sup>2</sup> MIT, [jvielma@mit.edu](mailto:jvielma@mit.edu)

**Abstract.** We give an explicit geometric way to build mixed-integer programming (MIP) formulations for unions of polyhedra. The construction is simply described in terms of spanning hyperplanes in an  $r$ -dimensional linear space. The resulting MIP formulation is ideal, and uses exactly  $r$  integer variables and  $2 \times (\#$  of spanning hyperplanes) general inequality constraints. We use this result to derive novel logarithmic-sized ideal MIP formulations for discontinuous piecewise linear functions and structures appearing in robotics and power systems problems.

**Keywords:** Mixed-integer programming, Formulations, Disjunctive constraints

## 1 Introduction

Consider a disjunctive set  $\bigcup_{i=1}^d S^i$ , where each *alternative*  $S^i \subset \mathbb{R}^n$  is a rational polyhedra. Disjunctive constraints of the form  $x \in \bigcup_{i=1}^d S^i$  abound in optimization: they are useful, for example, to model nonlinearities [1,2] or discrete logic imposed by complex processes [3,4]. Therefore, we would like a way to represent these constraints in such a way that we can efficiently optimize over them. Additionally, we would like to do this in a composable way, as disjunctive constraints frequently arise as substructures in large, complex optimization problems.

Mixed-integer programming (MIP) offers one such solution. MIP formulations are useful because there are sophisticated algorithms—and corresponding high-quality software implementations—that can optimize over these representations efficiently in practice [5,6]. Furthermore, combining MIP formulations for different substructures is trivial, and so this technology can be marshalled for very complex and large-scale optimization problems.

Our contribution in this work is an *explicit geometric construction for strong MIP formulations of disjunctive sets*. In particular, we give a constructive geometric description for the convex hull of a particularly structured set corresponding to the “embedding” of a disjunctive set in a higher-dimensional space. By carefully choosing how we do this embedding, we will be able to apply the main technical result to construct small, strong MIP formulations for certain disjunctive constraints of broad interest.

## 2 Preliminaries

Although predicting the relative performance of different MIP formulations is a difficult task, two properties that tend to correlate strongly with good computational performance are *strength* and *small size*. A MIP formulation for  $x \in \bigcup_{i=1}^d S^i \subseteq \mathbb{R}^n$  is given by its LP relaxation  $R \subseteq \mathbb{R}^{n+r}$ , where  $\text{Proj}_x(R \cap (\mathbb{R}^n \times \mathbb{Z}^r)) = \bigcup_{i=1}^d S^i$  is the orthogonal projection onto the  $x$  variables. In this work, we focus exclusively on *ideal formulations*, where the extreme points of the LP relaxation are naturally integral:  $\text{ext}(R) \subseteq \mathbb{R}^n \times \mathbb{Z}^r$ . Ideal formulations are the strongest possible formulations in terms of the LP relaxation, and tend to perform very well computationally (see [7] for a more detailed discussion). Additionally, we will seek formulations that are small, requiring few additional variables and constraints. More concretely, we will endeavor to build formulations with few integer variables  $r$  (ideally, logarithmic in  $d$ ), and few *general inequality* constraints (i.e. those not equivalent to variable bounds). We focus on general inequality constraints with since modern MIP solvers can incorporate variable bounds directly with very little overhead cost.

### 2.1 Combinatorial disjunctive constraints

We will focus on the subclass of disjunctive constraints known as *combinatorial disjunctive constraints* [8]. In particular, we assume that each alternative is a face of the unit simplex  $\Delta^V = \{ \lambda \in \mathbb{R}_{\geq 0}^V \mid \sum_{v \in V} \lambda_v = 1 \}$ , where  $\mathbb{R}_{\geq 0}^V$  is the nonnegative orthant in  $\mathbb{R}^V$ . Formally, this means we consider disjunctive constraints of the form  $\bigcup_{i=1}^d P(T^i)$  for some finite sets  $T^i \subseteq V$ , where  $P(T) \stackrel{\text{def}}{=} \{ \lambda \in \Delta^V \mid \lambda_v \neq 0 \text{ only if } v \in T \}$ . We will use  $\mathcal{T} = (T^i)_{i=1}^d$  notationally as the collection of subsets corresponding to each alternative.

This class of disjunctive constraints is far more expressive than it may appear at first glance. Suppose you have an arbitrary disjunctive constraint  $\bigcup_{i=1}^d S^i$ , presuming for the moment that each  $S^i$  is bounded. If you take each  $T^i = \text{ext}(S^i)$  as the set of extreme points of  $S^i$ , then

$$\bigcup_{i=1}^d S^i = \left\{ \sum_{v \in V} \lambda_v v \mid \exists \lambda \in \bigcup_{i=1}^d P(T^i) \right\}.$$

In other words, we can adapt a formulation for  $\bigcup_{i=1}^d P(T^i)$  to one for  $\bigcup_{i=1}^d S^i$  by introducing auxiliary convex multiplier variables  $\lambda \in \Delta^V$ . This readily generalizes to the case where the alternatives  $S^i$  are unbounded (provided standard representability conditions hold [7]); see Section 4.1 for an example.

### 2.2 Existing techniques

There are two standard formulation techniques which are generic in the sense that they can be applied to any combinatorial disjunctive constraint. The first is the well-known big- $M$  formulation approach; see, for example, [7, Corollary 6.3]. Big- $M$  formulations are simple to reason about and small, though they will, in

general, fail to be ideal. A second approach applies the techniques of Jeroslow and Lowe [9] to construct ideal *extended* formulations, requiring additional continuous variables: in particular, a copy of the  $\lambda_v$  variables for each alternative  $S^i$  and each  $v \in S^i$  [7, Corollary 5.2]. Despite the modest increase in size of the formulation, it has been observed that the resulting formulation can perform worse than expected, potentially due to the block structure of the constraints [10].

Alternatively, there exist a number of formulation techniques which are structure-dependent. While not applicable for every combinatorial disjunctive constraint, they can often be deployed to construct very small ideal formulations, the canonical example being the so-called “logarithmic” MIP formulations for the special ordered set of type 2 constraint [11], which have proven computational efficacy [7,12,13,14]. One example of a structure-dependent technique is a combinatorial approach previously proposed by the authors which is applicable for instances where a graph associated with the extreme points of the alternatives satisfies a certain condition [8]. The size of the resulting formulation is dependent on the size of a decomposition of this graph. Alternatively, there exists a geometric technique that characterizes all possible non-extended ideal formulations for the particular case of the SOS2 constraint [12,15]. The technique produces a formulation whose size is determined by the number of spanning hyperplanes induced by a particular set of vectors.

This paper builds on this geometric approach by greatly expanding its applicability. In particular, we present a technique to construct ideal MIP formulations, based on hyperplane arrangements, that can be applied to any combinatorial disjunctive constraint satisfying a mild connectivity condition.

### 2.3 The embedding approach

We construct formulations for combinatorial disjunctive constraints through the *embedding approach* [15], which works as follows. Assign each alternative  $P(T^i)$  a unique vector  $h^i \in \mathbb{R}^r$ . We call such a collection of distinct vectors  $H = (h^i)_{i=1}^d$  an *encoding*. Given  $\mathcal{T}$  and  $H$ , embed the disjunctive constraint in a higher dimensional space as:

$$\text{Em}(\mathcal{T}, H) \stackrel{\text{def}}{=} \bigcup_{i=1}^d (P(T^i) \times \{h^i\}).$$

If a natural geometric condition is satisfied, then the convex hull of the embedding,  $Q(\mathcal{T}, H) \stackrel{\text{def}}{=} \text{Conv}(\text{Em}(\mathcal{T}, H))$ , immediately gives an ideal formulation.

**Definition 1.** *A set  $H \subset \mathbb{R}^r$  is in convex position if  $\text{ext}(\text{Conv}(H)) = H$ , it is hole-free if  $\text{Conv}(H) \cap \mathbb{Z}^r = H$ .*

**Proposition 1.** *If  $H$  is hole-free and in convex position, then  $Q(\mathcal{T}, H)$  is the LP relaxation for an ideal formulation of  $\bigcup_{i=1}^d P(T^i)$ .*

*Proof.* Follows as a straightforward extension of [15, Proposition 1].

We present two encodings that we will make repeated use of later.

**Definition 2.** Take the families of matrices  $(K^s \in \{0, 1\}^{2^s \times s})_{s \in \mathbb{N}}$  and  $(C^s \in \mathbb{Z}^{2^s \times s})_{s \in \mathbb{N}}$  defined recursively, where  $K^1 = C^1 = (0, 1)^T$  and for each  $s \in \mathbb{N}$ ,

$$K^{s+1} = \begin{pmatrix} K^s & \mathbf{0}^{2^s} \\ \text{rev}(K^s) & \mathbf{1}^{2^s} \end{pmatrix} \quad \text{and} \quad C^{s+1} = \begin{pmatrix} C^s & \mathbf{0}^{2^s} \\ C^s + \mathbf{1}^{2^s} \otimes C_{2^s}^s & \mathbf{1}^{2^s} \end{pmatrix},$$

where  $\mathbf{0}^t, \mathbf{1}^t \in \{0, 1\}^t$  are  $t$ -dimensional vectors of all zeros and ones, respectively,  $\text{rev}(A)$  reverses the rows of the matrix  $A$ ,  $u \otimes v = uv^T \in \mathbb{R}^{m \times n}$  for any  $u \in \mathbb{R}^m$  and  $v \in \mathbb{R}^n$ , and  $A_i$  is the  $i$ -th row of matrix  $A$ .

For  $d \in \mathbb{N}$  and  $r = \lceil \log_2(d) \rceil$ , the binary reflected Gray encoding  $H_d^{\text{br}} \stackrel{\text{def}}{=} (K_i^r)_{i=1}^d$  and the zig-zag encoding  $H_d^{\text{zz}} \stackrel{\text{def}}{=} (C_i^r)_{i=1}^d$  are given by the sequence of the first  $d$  rows of  $K^r$  and  $C^r$ , respectively.

As a concrete example,

$$K^3 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad C^3 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 1 & 0 \\ 2 & 1 & 1 \\ 3 & 1 & 1 \\ 3 & 2 & 1 \\ 4 & 2 & 1 \end{pmatrix}.$$

### 3 The main result

To use an embedding formulation in practice, we need an explicit outer (inequality) description of  $Q(\mathcal{T}, H)$ . Our main technical result provides just such a description, hinging on the computation of a particular set of spanning hyperplanes. This result generalizes those of Huchette and Vielma [15, 12], which only apply to the special case where  $\mathcal{T} = (\{i, i+1\})_{i=1}^d$  (i.e. the SOS2 constraint of Beale and Tomlin [11]).

To simplify our discussion, we will assume without loss of generality (w.l.o.g.) that  $V = \llbracket n \rrbracket \stackrel{\text{def}}{=} \{1, 2, \dots, n\}$ .

**Theorem 1.** Given the family of sets  $\mathcal{T} = (T^i \subseteq \llbracket n \rrbracket)_{i=1}^d$  and an encoding  $H = (h^i)_{i=1}^d \subset \mathbb{R}^r$  in convex position, presume that  $\llbracket n \rrbracket = \bigcup_{i=1}^d T^i$  and take:

- the intersection directed graph  $D = \left\{ (i, j) \in \llbracket d \rrbracket^2 \mid T^i \cap T^j \neq \emptyset, i < j \right\}$ ,
- the set of difference directions  $C = \{c^{i,j} \stackrel{\text{def}}{=} h^j - h^i\}_{(i,j) \in D}$ ,
- the ambient linear space  $\mathcal{L} = \text{span}(C)$ ,
- the hyperplane defined by  $b$  in  $\mathcal{L}$ :  $M(b; \mathcal{L}) \stackrel{\text{def}}{=} \{y \in \mathcal{L} \mid b \cdot y = 0\}$ , and
- the normal directions  $\{b^k\}_{k=1}^r \subset \mathbb{R}^r \setminus \{\mathbf{0}^r\}$  for the hyperplanes  $\{M(b^k; \mathcal{L})\}_{k=1}^r$  spanned by  $C$  in  $\mathcal{L}$ .

If  $\dim(\text{span}(C)) = \dim(H)$ , then  $(\lambda, z) \in Q(\mathcal{T}, H)$  if and only if

$$\sum_{v=1}^n \min_{s:v \in T^s} \{b^k \cdot h^s\} \lambda_v \leq b^k \cdot z \leq \sum_{v=1}^n \max_{s:v \in T^s} \{b^k \cdot h^s\} \lambda_v \quad \forall k \in \llbracket \Gamma \rrbracket \quad (1a)$$

$$(\lambda, z) \in \Delta^n \times \text{aff}(H). \quad (1b)$$

We defer the proof of the result to Appendix A. We can present a straightforward sufficient condition that ensures that the dimensionality conditions of Theorem 1 are satisfied. Notationally, take  $[r]^2 \stackrel{\text{def}}{=} \{ \{i, j\} \mid i, j \in \llbracket r \rrbracket, i \neq j \}$ .

**Proposition 2.** *If the directed graph  $G = (\llbracket d \rrbracket, D)$  is (weakly) connected, then  $\dim(\text{span}(C)) = \dim(H)$ .*

We defer the proof of the result to Appendix B.

## 4 Applications of the main result

### 4.1 Univariate piecewise linear functions (continuous or discontinuous)

We can apply Theorem 1 to construct small, strong MIP formulations for univariate piecewise linear functions that have a sufficiently long continuous sub-piece. This application includes as special cases the existing best-of-breed MIP formulations for continuous piecewise linear functions (the “logarithmic” formulation of Vielma and Nemhauser [15,14] and the “zig-zag” formulation of Huchette and Vielma [12]), and improves upon standard formulations described in [13] for discontinuous piecewise linear functions.

Consider a lower semi-continuous univariate piecewise linear function

$$f(x) = \begin{cases} a_1 x + b_1 & t_1 \leq x \leq t_2 \\ \vdots & \\ a_d x + b_d & t_d < x \leq t_{d+1}. \end{cases}$$

We can model its epigraph as a union of polyhedra:

$$\text{epi}(f) = \bigcup_{i=1}^d \{ (x, a_i x + b_i + \gamma) \mid t_i \leq x \leq t_{i+1}, \gamma \geq 0 \}.$$

To formulate this set, [13] duplicates the  $\lambda$  variables for each interior breakpoint  $\{t_i\}_{i=2}^d$ . If  $r(i) = \lfloor 1 + i/2 \rfloor$  and  $s(i) = \lceil i/2 \rceil$ , this takes the form

$$\text{epi}(f) = \sum_{i=1}^{2d} \lambda_i (t_{r(i)}, a_{s(i)} t_{r(i)} + b_{s(i)} + \gamma), \quad \gamma \geq 0, \lambda \in \Delta^{2d}, \lambda \text{ is SOS2}, \quad (2)$$

where we have constrained  $\lambda$  to satisfy the standard SOS2 constraint on  $2d$  breakpoints. However, this approach is inefficient as it unnecessarily duplicates breakpoints where  $f$  is continuous (i.e. those  $i$  where  $a_i t_{i+1} + b_i = a_{i+1} t_{i+1} + b_{i+1}$ ). We can remove this redundancy by considering the family of sets

$$T^i \stackrel{\text{def}}{=} \{(t_i, a_i t_i + b_i), (t_{i+1}, a_i t_{i+1} + b_i)\} \quad \forall i \in \llbracket d \rrbracket.$$

We can then identify the ground set  $\bigcup_{i=1}^d T^i$  with  $V = \llbracket d + 1 + \kappa \rrbracket$ , where  $\kappa$  is the number of discontinuous breakpoints, i.e. those  $i \in \llbracket d \rrbracket$  where  $a_i t_{i+1} + b_i \neq a_{i+1} t_{i+1} + b_{i+1}$ . The mapping between  $\bigcup_{i=1}^d T^i$  and  $V$  is the natural extension of that in (2) where all breakpoints are duplicated. Provided that  $f$  has a sufficiently long continuous sub-piece, we can construct a logarithmically-sized ideal MIP formulation for  $\text{epi}(f)$ .

**Proposition 3.** *Take  $d = 2^r$  for some integer  $r \geq 2$ , and assume that  $f$  is continuous on the interval  $[t_{d/4+1}, t_{d/2+1}]$ , inclusive of the endpoints. Select an encoding  $H = (h^i)_{i=1}^d$  that is either  $H_d^{\text{br}}$  or  $H_d^{\text{zz}}$ , and take  $h^0 \equiv h^1$  and  $h^{d+1} \equiv h^d$  for notational simplicity. Then  $(\lambda, z) \in Q(\mathcal{T}, H)$  if and only if*

$$\sum_{v=1}^{d+1+\kappa} \lambda_v \min_{s:v \in T^s} h_k^s \leq z_k \leq \sum_{v=1}^{d+1+\kappa} \lambda_v \max_{s:v \in T^s} h_k^s \quad \forall k \in \llbracket r \rrbracket \quad (3a)$$

$$(\lambda, z) \in \Delta^{d+1+\kappa} \times \mathbb{R}^r. \quad (3b)$$

Moreover, a directly analogous result can be obtained with continuity on the interval  $[t_{d/2+1}, t_{3d/4+1}]$ .

*Proof.* The result follows by inspection of the recursive definitions of  $H_d^{\text{br}}$  and  $H_d^{\text{zz}}$ , seeing that under the partial continuity assumption that  $C \supseteq \{\mathbf{e}^k\}_{k=1}^r$ , and so  $C$  has dimension  $r$ . Furthermore, as  $C \subseteq \{\pm \mathbf{e}^k\}_{k=1}^r$ , we can apply Theorem 1 using the normal directions  $\{b^k = \mathbf{e}^k\}_{k=1}^r$ , giving the result.

Applying one of the existing logarithmic formulations for the SOS2 constraint in (2) yields a formulation with the same number of binary variables and general inequalities as (3). However, the formulation based on (2) will have  $2d$  continuous variables even if the number of discontinuous breakpoints  $\kappa$  is small or even zero. In contrast, we observe that formulation (3) uses only  $d + 1 + \kappa$  continuous variables. A similar favorable accounting occurs when comparing (3) with the “disaggregated logarithmic” formulation described in [13]. Additionally, when  $\kappa = 0$ , (3) is equivalent to the logarithmic formulations for continuous functions from [12,14,15].

## 4.2 Relaxations of the annulus

The annulus is a set in the plane  $\mathcal{A} = \{x \in \mathbb{R}^2 \mid L \leq \|x\|_2 \leq U\}$  for constants  $L, U \in \mathbb{R}_{\geq 0}$ ; see the left side of Figure 1 for an illustration. A constraint of the form  $x \in \mathcal{A}$  might arise when modeling a complex number  $z = x_1 + x_2 \mathbf{i}$ , as  $x \in \mathcal{A}$  bounds the magnitude of  $z$  as  $L \leq |z| \leq U$ . Such constraints arise in power systems optimization: for example, in the “rectangular formulation” [16] and the second-order cone reformulation [17,18] of the optimal power flow problem and its voltage stability-constrained variant [19], and the reactive power dispatch problem [20]. Another application is footstep planning in robotics [21,22], where  $L = U = 1$ ,  $x = (\cos(\theta), \sin(\theta))$ , and  $x$  must satisfy the trigonometric identity  $x_1^2 + x_2^2 = 1$ .

When  $0 < L \leq U$ ,  $\mathcal{A}$  is a nonconvex set. Moreover, the annulus is not *mixed-integer convex representable* [23,24]: that is, there do not exist mixed-integer formulations for the annulus even if we allow the relaxation  $R$  to be an arbitrary convex set. Foster [20] proposes a disjunctive relaxation for the annulus given as  $\hat{\mathcal{A}} \stackrel{\text{def}}{=} \bigcup_{i=1}^d S^i$ , where each

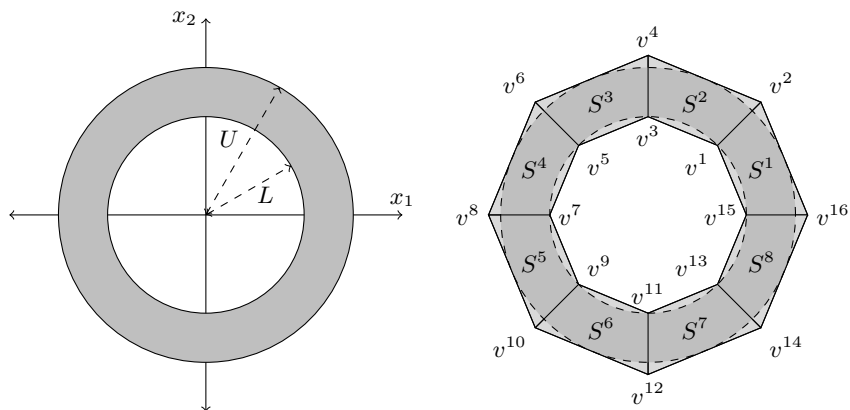
$$S^i = \text{Conv} \left( \left\{ v^{2i+s-4} \right\}_{s=1}^4 \right) \quad \forall i \in \llbracket d \rrbracket \tag{4}$$

is a quadrilateral whose extreme points are

$$\begin{aligned} v^{2i-1} &= \left( L \cos \left( \frac{2\pi i}{d} \right), L \sin \left( \frac{2\pi i}{d} \right) \right) & \forall i \in \llbracket d \rrbracket \\ v^{2i} &= \left( U \sec \left( \frac{2\pi}{d} \right) \cos \left( \frac{2\pi i}{d} \right), U \sec \left( \frac{2\pi}{d} \right) \sin \left( \frac{2\pi i}{d} \right) \right) & \forall i \in \llbracket d \rrbracket, \end{aligned}$$

taking  $v^0 \equiv v^{2d}$  and  $v^{-1} \equiv v^{2d-1}$  for notational simplicity. We can in turn represent this disjunctive relaxation via the family of sets  $\mathcal{T} = (T^i = \{2i + s - 4\}_{s=1}^4)_{i=1}^d$ ; see the right side of Figure 1 for an illustration.

We start by applying Theorem 1 to construct an ideal MIP formulation with  $\lceil \log_2(d) \rceil$  integer variables and  $2\lceil \log_2(d) \rceil$  inequality constraints.



**Fig. 1.** (Left) The annulus  $\mathcal{A}$  and (Right) its corresponding quadrilateral relaxation  $\hat{\mathcal{A}}$  given by (4) with  $d = 8$ .

**Proposition 4.** Fix  $d = 2^r$  for some  $r \in \mathbb{N}$ . Take the binary reflected Gray encoding  $H_d^{\text{br}} = (h^i)_{i=1}^d$ , along with  $h^0 \equiv h^d$  for notational convenience. Then

$(\lambda, z) \in Q(\mathcal{T}, H_d^{\text{br}})$  if and only if

$$\sum_{i=1}^d \min\{h_k^{i-1}, h_k^i\}(\lambda_{2i-1} + \lambda_{2i}) \leq z_k \leq \sum_{i=1}^d \max\{h_k^{i-1}, h_k^i\}(\lambda_{2i-1} + \lambda_{2i}) \quad \forall k \in \llbracket r \rrbracket \quad (5a)$$

$$(\lambda, z) \in \Delta^{2d} \times \mathbb{R}^r. \quad (5b)$$

*Proof.* The result follows from Theorem 1 after observing that  $D = \{(i, i+1)\}_{i=1}^{d-1} \cup (1, d)$  and therefore that  $C = \{\pm \mathbf{e}^k\}_{k=1}^r$ , as the binary reflected Gray encoding is cyclic (i.e.  $h^d - h^1 = \mathbf{e}^1$ ).

We can also apply Theorem 1 using the zig-zag encoding to produce a MIP formulation for the annulus with  $\lceil \log_2(d) \rceil$  integer variables and  $\mathcal{O}(\log^2(d))$  general inequality constraints. Despite the modest increase in the number of constraints, the analysis of Huchette and Vielma [12] shows that the zig-zag encoding enjoys the “incremental branching” behavior for univariate piecewise linear functions, leading to improved computational performance relative to the logarithmic formulation of Vielma et al. [13,14]. Therefore, it may be the case that the zig-zag formulation for the annulus similarly outperforms the logarithmic formulation (5), despite the modest increase in the number of constraints.

**Proposition 5.** *Fix  $d = 2^r$  for some  $r \in \mathbb{N}$ . Take the zig-zag encoding  $H_d^{\text{zz}} = (h^i)_{i=1}^d$ , along with  $h^0 \equiv h^d$  for notational convenience. Then  $(\lambda, z) \in Q(\mathcal{T}, H_d^{\text{zz}})$  if and only if*

$$\begin{aligned} \sum_{i=1}^d \min\{h_k^{i-1}, h_k^i\}(\lambda_{2i-1} + \lambda_{2i}) &\leq z_k \quad \forall k \in \llbracket r \rrbracket \\ \sum_{i=1}^d \max\{h_k^{i-1}, h_k^i\}(\lambda_{2i-1} + \lambda_{2i}) &\geq z_k \quad \forall k \in \llbracket r \rrbracket \\ \sum_{i=1}^d \min\left\{\frac{h_k^{i-1}}{2^\ell} - \frac{h_\ell^{i-1}}{2^k}, \frac{h_k^i}{2^\ell} - \frac{h_\ell^i}{2^k}\right\}(\lambda_{2i-1} + \lambda_{2i}) &\leq \frac{z_k}{2^\ell} - \frac{z_\ell}{2^k} \quad \forall \{k, \ell\} \in \llbracket r \rrbracket^2 \\ \sum_{i=1}^d \max\left\{\frac{h_k^{i-1}}{2^\ell} - \frac{h_\ell^{i-1}}{2^k}, \frac{h_k^i}{2^\ell} - \frac{h_\ell^i}{2^k}\right\}(\lambda_{2i-1} + \lambda_{2i}) &\geq \frac{z_k}{2^\ell} - \frac{z_\ell}{2^k} \quad \forall \{k, \ell\} \in \llbracket r \rrbracket^2 \\ (\lambda, z) &\in \Delta^{2d} \times \mathbb{R}^r. \end{aligned}$$

*Proof.* The result follows from Theorem 1. As  $D = \{(i, i+1)\}_{i=1}^{d-1} \cup (1, d)$ , it follows that  $C = \{\mathbf{e}^k\}_{k=1}^r \cup \{w \equiv (2^{r-1}, 2^{r-2}, \dots, 2^0)\}$ . We have that  $B = \{\mathbf{e}^k\}_{k=1}^r$  induce all hyperplanes spanned by the vectors  $C \setminus \{w\} = \{\mathbf{e}^k\}_{k=1}^r$ . Now consider each hyperplane spanned by  $\hat{C} = \{\mathbf{e}^k\}_{k \in I} \cup \{w\} \subset C$ , where  $I \subseteq \llbracket r \rrbracket$ . As  $|C| = r+1$  and  $\dim(\text{span}(C)) = r$ , we must have  $|I| = r-2$ , i.e. that there are distinct indices  $k, \ell \in \llbracket r \rrbracket \setminus I$  where  $I \cup \{k, \ell\} = \llbracket r \rrbracket$ . We may then compute that the corresponding hyperplane is given by the normal direction  $b^{k,\ell} \stackrel{\text{def}}{=} 2^{-\ell} \mathbf{e}^k - 2^{-k} \mathbf{e}^\ell$ . Therefore, we have that the set  $B = \{\mathbf{e}^k\}_{k=1}^r \cup \{b^{k,\ell}\}_{\{k,\ell\} \in \llbracket r \rrbracket^2}$  suffices for the conditions of Theorem 1, giving the result.



## A Proof of Theorem 1

*Proof.* We start by taking  $B = \text{ext}(Q(\mathcal{T}, H)) = \{(\mathbf{e}^w, h^j) \mid w \in T^j\}$  as the set of all extreme points of  $Q(\mathcal{T}, H)$ . It is straightforward to show that each  $(\lambda, z) \in Q(\mathcal{T}, H)$  satisfies (1). Take some  $(\hat{\lambda}, \hat{z}) = (\mathbf{e}^w, h^j) \in B$ . Clearly (1b) is satisfied, and we have that for each  $k \in \llbracket I \rrbracket$ ,

$$\sum_{v=1}^n \min_{s:w \in T^s} \{b^k \cdot h^s\} \hat{\lambda}_v = \min_{s:w \in T^s} \{b^k \cdot h^s\} \leq b^k \cdot h^j = b^k \cdot \hat{z},$$

where the inequality follows as  $w \in T^j$ . An identical argument follows for the other inequality in (1a). This implies that each  $(\lambda, z) \in Q(\mathcal{T}, H)$  satisfies (1), giving one direction of the characterization.

For the other direction, let  $F$  be a facet of  $Q(\mathcal{T}, H)$ . By possibly adding or subtracting multiples of  $\sum_{i=1}^n \lambda_i = 1$  (an equation appearing in the definition of  $\Delta^n$ ) and the equations defining  $\text{aff}(H)$ , we may assume w.l.o.g. that  $F$  is induced by  $\tilde{a} \cdot \lambda \leq \tilde{b} \cdot y$  for some  $(\tilde{a}, \tilde{b}) \in \mathbb{R}^{n+r}$ . As  $F$  is a facet, it is supported by some strict nonempty subset of extreme points  $\tilde{B} \subsetneq B$ . Take  $\tilde{D} = \{(i, j) \in D \mid \exists v \in \llbracket n \rrbracket \text{ s.t. } (\mathbf{e}^v, h^i), (\mathbf{e}^v, h^j) \in \tilde{B}\}$  and  $\tilde{C} = \{c^{i,j} \in C \mid (i, j) \in \tilde{D}\}$ . In particular, we see that  $\tilde{b} \cdot c^{i,j} = 0$  for each  $c^{i,j} \in \tilde{C}$ , as if  $(i, j) \in \tilde{D}$ , this implies that there is some  $v \in \llbracket n \rrbracket$  whereby  $\tilde{a} \cdot \mathbf{e}^v = \tilde{b} \cdot h^i = \tilde{b} \cdot h^j$ .

As  $F$  is a proper face, we know that there is some point in  $B$  not supporting  $F$ , w.l.o.g.  $(\mathbf{e}^1, h^1) \in B \setminus \tilde{B}$ . We will take all the remaining extreme points not in  $\tilde{B}$  as  $B^* = B \setminus (\tilde{B} \cup \{(\mathbf{e}^1, h^1)\})$ . Consider three cases for the dimension of  $\text{span}(\tilde{C})$  which exhaust all possibilities.

Case 1:  $\dim(\text{span}(\tilde{C})) = \dim(\text{span}(C))$  In this case, we show that  $F$  corresponds to a variable bound on a single component of  $\lambda$ . As  $\tilde{C} \subseteq C$  and  $\dim(\text{span}(\tilde{C})) = \dim(\text{span}(C))$ , we conclude that  $\text{span}(\tilde{C}) = \text{span}(C) \equiv \mathcal{L}$ . Then  $\tilde{b} \in \mathcal{L}^\perp$ , as  $\tilde{b} \perp \tilde{C}$ . Furthermore,  $\mathcal{L}$  is the linear space parallel to  $\text{aff}(H)$ . Therefore, we can w.l.o.g. presume that  $\tilde{b} = \mathbf{0}^r$ , as (1b) constrains  $z \in \text{aff}(H)$ .

We observe that  $\tilde{a} \neq \mathbf{0}^n$ , as otherwise this would correspond to the vacuous inequality  $0 \leq 0$ , which is not a proper face. We now show that  $\tilde{a}$  has exactly one nonzero element. Assume for contradiction that this is not the case, and  $\tilde{a}_p, \tilde{a}_q < 0$  for some distinct  $p, q \in \llbracket n \rrbracket$  (any strictly positive components will not yield a valid inequality for  $B$ ). This implies that there exists some  $j, k \in \llbracket d \rrbracket$  such that at least one of  $(\mathbf{e}^p, h^j)$  and  $(\mathbf{e}^q, h^k)$  is not contained in  $B^*$ ; assume w.l.o.g. that  $(\mathbf{e}^p, h^j) \notin B^*$ . However, we could then perform the simple tilting  $\tilde{a}_q \leftarrow 0$  to construct a distinct face with strictly larger support, as now  $(\mathbf{e}^q, h^j)$  is supported by the corresponding face for each  $j$  such that  $q \in T^j$ . Furthermore, as this new constraint does not support  $(\mathbf{e}^p, h^j)$  for each  $j$  such that  $p \in T^j$ , the new face is proper, and thus contradicts the original face  $F$  being a facet. Therefore, we can normalize the coefficients to  $\tilde{a} = -\mathbf{e}^p$ , giving a variable bound constraint on a component of  $\lambda$  which appears in the restriction  $\lambda \in \Delta^n$  in (1b).

Case 2:  $\dim(\text{span}(\tilde{C})) = \dim(\text{span}(C)) - 1$  The fact that  $\tilde{b} \perp \tilde{C}$ , along with the dimensionality of  $\tilde{C}$ , implies that  $M(\tilde{b}; \mathcal{L}) = \text{span}(\tilde{C})$  is a hyperplane in  $\mathcal{L}$ . This means we can assume w.l.o.g. that  $\tilde{b} = sb^k$  for some  $k \in \llbracket T \rrbracket$  and  $s \in \{-1, +1\}$ . We then compute for each  $v \in \llbracket n \rrbracket$  that either  $a_v = \min_{j:v \in T^j} \{b^k \cdot h^j\}$  if  $s = +1$ , or  $a_v = -\max_{j:v \in T^j} \{b^k \cdot h^j\}$  if  $s = -1$ . Therefore, this facet is included in (1a).



**Fig. 2.** (Left)  $\dim(L) = 1$ , and so we cannot “tilt” the inequality to increase its support. (Right)  $\dim(\text{span}(C)) = \dim(H)$  implies  $\dim(L) > 1$  (i.e. all of  $\mathbb{R}^2$  in this projected space), and so we can tilt.

Case 3:  $\dim(\text{span}(\tilde{C})) < \dim(\text{span}(C)) - 1$  We will show that this case cannot occur if  $F$  is a general inequality facet. In fact, observe that if  $\mathbf{e}^i \notin \text{Proj}_\lambda(\tilde{B})$ , then  $\tilde{a} \cdot \lambda \leq \tilde{b} \cdot z$  is either equivalent to, or dominated by, the variable bound  $\lambda_i \geq 0$ . Therefore, we assume that  $\text{Proj}_\lambda(\tilde{B}) = \{\mathbf{e}^i\}_{i=1}^n$  for the remainder.

Presume for contradiction that it is indeed the case that  $F$  is a facet and  $\dim(\text{span}(\tilde{C})) < \dim(\text{span}(C)) - 1$ . First, we show that  $B^* \neq \emptyset$ . If this were not the case, then  $\tilde{B} = B \setminus \{(\mathbf{e}^1, h^1)\}$ , and hence for each  $(i, j) \in D \setminus \tilde{D}$ , it must be that  $i = 1$  and  $T^1 \cap T^j = \{1\}$ . Then there exists  $J \subseteq \{2, \dots, d\}$  such that  $\tilde{C} = C \setminus \{c^{1,j} \mid j \in J\}$  and  $C_J \stackrel{\text{def}}{=} \{c^{i,j} \mid i, j \in J, i < j\} \subseteq C$ . Without loss of generality we may assume that  $J = \{2, \dots, k\}$ . Then by noting that  $c^{1,j} \equiv h^j - h^1 = \sum_{i=2}^j c^{i-1,i} = \sum_{i=2}^j h^i - h^{i-1}$ , we conclude that  $\dim(C_J \setminus \{c^{1,j} \mid j \in J\}) \geq \dim(C_J) - 1$ , which contradicts  $\dim(\text{span}(\tilde{C})) < \dim(\text{span}(C)) - 1$ . Therefore, we conclude that  $B^* \neq \emptyset$ .

We now define the cone  $K = \{(a, b) \in \mathbb{R}^n \times \mathcal{L} \mid a \cdot \mathbf{e}^v \leq b \cdot h^j \forall (\mathbf{e}^v, h^j) \in B^*\}$  and the linear space  $L = \{(a, b) \in \mathbb{R}^n \times \mathcal{L} \mid a \cdot \mathbf{e}^v = b \cdot h^j \forall (\mathbf{e}^v, h^j) \in \tilde{B}\}$ .

By definition of  $B^*$ ,  $(\tilde{a}, \tilde{b}) \in \text{reint}(K)$  (taken relative to  $\mathbb{R}^n \times \mathcal{L}$ ). Next, we show that  $\dim(L) > 1$ . To show this, we start by instead studying  $L' = \{b \in \mathcal{L} \mid b \cdot c = 0 \forall c \in \tilde{C}\}$ . We can readily observe that  $L' = \text{Proj}_b(L)$ . Furthermore, as  $\text{Proj}_a(\tilde{B}) = \{\mathbf{e}^i\}_{i=1}^n$  from the argument at the beginning of the case, we conclude that the set  $\{a \mid (a, b) \in L\}$  is a singleton. In other words, the values for  $a$  are completely determined by the values for  $b$  in  $L$ . From this, we conclude that  $\dim(L) = \dim(L')$ . From the definition of  $L'$ , we see that

$L'$  and  $\text{span}(\tilde{C})$  form an orthogonal decomposition of  $\mathcal{L}$ . Therefore,  $\dim(\mathcal{L}) = \dim(L') + \dim(\text{span}(\tilde{C}))$ . Recalling that  $\dim(\mathcal{L}) = \dim(\text{span}(C))$ , and that we are assuming that  $\dim(\text{span}(\tilde{C})) < \dim(\text{span}(C)) - 1$ , we have that  $\dim(L) = \dim(L') = \dim(\mathcal{L}) - \dim(\text{span}(\tilde{C})) = \dim(\text{span}(C)) - \dim(\text{span}(\tilde{C})) > 1$ , giving the result.

We now show that  $K \cap L$  is pointed. To see this, presume for contradiction that there exists a nonzero  $(\hat{a}, \hat{b})$  such that  $(\hat{a}, \hat{b}), (-\hat{a}, -\hat{b}) \in K \cap L$ . However, this would imply that  $\hat{a} \cdot \mathbf{e}^v = \hat{b} \cdot h^j$  for all  $(\mathbf{e}^v, h^j) \in \tilde{B} \cup B^*$ . Because  $B^* \neq \emptyset$ , this implies that  $\hat{a} \cdot \lambda \leq \hat{b} \cdot z$  is a face strictly containing the facet  $F$ , and so must be a non-proper face (i.e. it is additionally supported by  $(\mathbf{e}^1, h^1)$  and hence by every point in  $B$ ). However, this would imply that  $\hat{b} \cdot c = 0$  for all  $c \in C$ , and as  $\mathcal{L} = \text{span}(C)$ , this would necessitate that  $\hat{b} \in \mathcal{L}^\perp$ . As  $\hat{b} \in \mathcal{L}$  from the definition of  $K$ , it follows that  $\hat{b} = \mathbf{0}^r$ . However, this would imply that  $\hat{a} \cdot \lambda = 0$  is valid for  $B$ , which cannot be the case unless  $\hat{a} = \mathbf{0}^n$ , a contradiction. Therefore,  $K \cap L$  is pointed.

As  $\dim(L) > 1$ , we can take some two-dimensional linear subspace  $L^2 \subseteq L$  such that  $(\tilde{a}, \tilde{b}) \in L^2$ . As  $(\tilde{a}, \tilde{b}) \in L \cap \text{relint}(K)$ , it follows that  $(\tilde{a}, \tilde{b}) \in L^2 \cap \text{relint}(K)$  as well. Similarly, as  $K \cap L$  is pointed, it follows that  $K^2 \stackrel{\text{def}}{=} L^2 \cap K$  is pointed as well. Furthermore, as  $K$  is full-dimensional in  $\mathbb{R}^n \times \mathcal{L}$ ,  $K^2$  is full-dimensional in  $L^2 \subset \mathbb{R}^n \times \mathcal{L}$  (i.e. 2-dimensional). Therefore, a minimal description for it includes the equalities that define  $L^2$ , along with exactly two nonempty-face-inducing inequality constraints from the definition of  $K$ . See Figure 2 for an illustration of the following argument in this space.

Add the single strict inequality  $\hat{K}^2 \stackrel{\text{def}}{=} K^2 \cap \{ (a, b) \in \mathbb{R}^n \times \mathcal{L} \mid a \cdot \mathbf{e}^1 < b \cdot h^1 \}$ . As  $\tilde{a} \cdot \mathbf{e}^1 < \tilde{b} \cdot h^1$  and  $(\tilde{a}, \tilde{b}) \in K^2$ , it follows that  $\hat{K}^2$  is nonempty and also 2-dimensional, and can be described using only the linear equations defining  $L^2$ , the strict inequality  $a \cdot \mathbf{e}^1 < b \cdot h^1$ , and at least one (and potentially two) of the inequalities previously used to describe  $K^2$ . Select one of the defining nonempty-face-inducing inequalities given by  $a \cdot \mathbf{e}^v \leq b \cdot h^j$ , where  $(\mathbf{e}^v, h^j) \in B^*$ .

Now construct the restriction  $S \stackrel{\text{def}}{=} \{ (a, b) \in \hat{K}^2 \mid a \cdot \mathbf{e}^v = b \cdot h^j \}$ . As  $a \cdot \mathbf{e}^v \leq b \cdot h^j$  induces a non-empty face on the cone  $\hat{K}^2$ ,  $S$  is nonempty. Furthermore, we see that any  $(\hat{a}, \hat{b}) \in S$  will correspond to a valid inequality  $\hat{a} \cdot \lambda \leq \hat{b} \cdot z$  for  $B$  with strictly greater support than our original face  $\tilde{a} \cdot \lambda \leq \tilde{b} \cdot z$ . In particular, we see that  $(\mathbf{e}^v, h^j) \in B^*$ , i.e.  $\tilde{a} \cdot \mathbf{e}^v < \tilde{b} \cdot h^j$ , but by construction  $\hat{a} \cdot \mathbf{e}^v = \hat{b} \cdot h^j$ . Additionally, since  $\hat{a} \cdot \mathbf{e}^1 < \hat{b} \cdot h^1$ , the corresponding face is proper, which implies that  $F$  cannot be a facet, a contradiction.

## B Proof of Proposition 2

*Proof.* The result follows by showing that  $\mathcal{L} = \text{aff}(H) - h^1$ . For simplicity, take  $\check{D} = \{ (i, j) \in [d]^2 \mid T^i \cap T^j \neq \emptyset \}$  as the version of  $D$  with all parallel edges added. To show that  $\text{span}(C) \subseteq \text{aff}(H) - h^1$ , take some  $z \in \text{span}(C)$ , and so there exist multipliers  $\gamma_{i,j}$  such that  $z = \sum_{(i,j) \in \check{D}} \gamma_{i,j} (h^i - h^j) = \sum_{i=1}^d \alpha_i h^i$ , where  $\alpha_i =$

$\sum_{j:(i,j) \in \check{D}} \gamma_{i,j} - \sum_{k:(k,i) \in \check{D}} \gamma_{k,i}$ . Then  $z = (\alpha_1 + 1)h^1 + (\sum_{i=2}^d \alpha_i h^i) - h^1$ , i.e.  $z \in \text{aff}(H) - h^1$ , as  $(\alpha_1 + 1) + \sum_{i=2}^d \alpha_i = 1 + \sum_{i=1}^d \alpha_i = 1 + \sum_{i=1}^d \left( \sum_{j:(i,j) \in \check{D}} \gamma_{i,j} - \sum_{k:(k,i) \in \check{D}} \gamma_{k,i} \right) = 1 + 0$ , and so they form affine multipliers.

To show that  $\text{span}(C) \supseteq \text{aff}(H) - h^1$ , take some  $z \in \text{aff}(H) - h^1$ , and so there exists multipliers  $\mu_i$  such that  $z = (\sum_{i=1}^d \mu_i h^i) - h^1$  and  $\sum_{i=1}^d \mu_i = 1$ . As  $G$  is connected, there exists some closed directed path  $(t_1 \equiv 1, t_2, \dots, t_r, t_{r+1} \equiv 1)$  on  $G$  that traverses each node. Take  $\alpha_i \stackrel{\text{def}}{=} -\frac{\mu_i}{\# \text{ of times path traverses } i}$  for each  $i \in \llbracket d \rrbracket$ . Then  $z = \sum_{k=1}^r (h^{t_{k+1}} - h^{t_k}) \sum_{\ell=1}^k \alpha_{t_\ell} = \sum_{k=1}^r c^{t_k, t_{k+1}} \sum_{\ell=1}^k \alpha_{t_\ell}$  (using  $\sum_{i=1}^d \mu_i = 1$  to show that the  $\ell = r$  term in the sum produces the desired  $-h^1$  term), and so therefore  $z \in \text{span}(C)$ , as each  $(t_k, t_{k+1}) \in \check{D}$ . This shows the result. Additionally, we note that the choice of  $h^1$  to subtract from  $\text{aff}(H)$  was arbitrary.

## References

1. B. Geißler, A. Martin, A. Morsi, L. Schewe, Using piecewise linear functions for solving MINLPs, Springer, 2012, pp. 287–314.
2. R. Misener, C. Floudas, Global optimization of mixed-integer quadratically-constrained quadratic programs (MIQCQP) through piecewise-linear and edge-concave relaxations, *Mathematical Programming* 136 (1) (2012) 155–182.
3. M. L. Bergamini, P. Aguirre, I. Grossmann, Logic-based outer approximation for globally optimal synthesis of process networks, *Computers and Chemical Engineering* 29 (9) (2005) 1914–1933.
4. F. Trespalcios, I. Grossmann, Review of mixed-integer nonlinear and generalized disjunctive programming methods, *Chemie Ingenieur Technik* 86 (7) (2014) 991–1012.
5. R. Bixby, E. Rothberg, Progress in computational mixed integer programming—A look back from the other side of the tipping point, *Annals of Operations Research* 149 (2007) 37–41.
6. M. Jünger, T. Liebling, D. Naddef, G. Nemhauser, W. Pulleyblank, G. Reinelt, G. Rinaldi, L. Wolsey, 50 years of integer programming 1958-2008, Springer, 2010.
7. J. P. Vielma, Mixed integer linear programming formulation techniques, *SIAM Review* 57 (1) (2015) 3–57.
8. J. Huchette, J. P. Vielma, A combinatorial approach for small and strong formulations of disjunctive constraints, *Mathematics of Operations Research* 44 (3) (2019) 767–1144.
9. R. Jeroslow, J. Lowe, Modelling with integer variables, *Mathematical Programming Study* 22 (1984) 167–184.
10. J. P. Vielma, Small and strong formulations for unions of convex sets from the Cayley embedding, *Mathematical Programming* (2018).
11. E. M. L. Beale, J. A. Tomlin, Special facilities in a general mathematical programming system for non-convex problems using ordered sets of variables, in: J. Lawrence (Ed.), *OR 69: Proceedings of the Fifth International Conference on Operational Research*, Tavistock Publications, 1970, pp. 447–454.
12. J. Huchette, J. P. Vielma, Nonconvex piecewise linear functions: Advanced formulations and simple modeling tools, <https://arxiv.org/abs/1708.00050> (2017).

13. J. P. Vielma, S. Ahmed, G. Nemhauser, Mixed-integer models for nonseparable piecewise-linear optimization: Unifying framework and extensions, *Operations Research* 58 (2) (2010) 303–315.
14. J. P. Vielma, G. Nemhauser, Modeling disjunctive constraints with a logarithmic number of binary variables and constraints, *Mathematical Programming* 128 (1-2) (2011) 49–72.
15. J. P. Vielma, Embedding formulations and complexity for unions of polyhedra, *Management Science* 64 (10) (2018) 4471–4965.
16. B. Kocuk, S. S. Dey, X. A. Sun, Strong SOCP relaxations for optimal power flow, *Operations Research* 64 (6) (2016) 1177–1196.
17. R. A. Jabr, Radial distribution load flow using conic programming, *IEEE Transactions on Power Systems* 21 (3) (2006) 1458–1459.
18. J. Liu, M. Bynum, A. Castillo, J.-P. Watson, C. D. Laird, A multitree approach for global solution of ACOPF problems using piecewise outer approximations, *Computers and Chemical Engineering* (2017).
19. B. Cui, X. A. Sun, A new voltage stability-constrained optimal power flow model: Sufficient condition, SOCP representation, and relaxation, *IEEE Transactions on Power Systems* (2017).
20. J. D. Foster, Mixed-integer quadratically-constrained programming, piecewise-linear approximation and error analysis with applications in power flow, Ph.D. thesis, The University of Newcastle, Australia (November 2013).
21. R. Deits, R. Tedrake, Footstep planning on uneven terrain with mixed-integer convex optimization, in: 2014 14th IEEE-RAS International Conference on Humanoid Robots (Humanoids), IEEE, 2014, pp. 279–286.
22. S. Kuindersma, R. Deits, M. Fallon, A. Valenzuela, H. Dai, F. Permenter, T. Koolen, P. Marion, R. Tedrake, Optimization-based locomotion planning, estimation, and control design for the atlas humanoid robot, *Autonomous Robots* 40 (3) (2016) 429–455.
23. M. Lubin, I. Zadik, J. P. Vielma, Mixed-integer convex representability, in: F. Eisenbrand (Ed.), *International Conference on Integer Programming and Combinatorial Optimization*, Springer International Publishing, Cham, 2017, pp. 392–404.
24. M. Lubin, I. Zadik, J. P. Vielma, Regularity in mixed-integer convex representability, <https://arxiv.org/abs/1706.05135> (2017).