Normalizations in Lift-and-Project Methods. A view from Convex Analysis

Pablo A. Rey [a,*] Claudia A. Sagastizábal [b,**]

[a] *PUC, Depto. Engenharia Elétrica. Rio de Janeiro, RJ, CEP 22453-900, Brazil.*
E-mail: prey@ele.puc-rio.br

[b] *IMPA, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil.*
On leave from INRIA-Rocquencourt, BP 105, 78153 Le Chesnay, France.
E-mail: sagastiz@impa.br

Branch-and-Cut algorithms for general 0-1 mixed integer programs can be successfully implemented by using Lift-and-Project (L&P) methods to generate cuts. L&P cuts are drawn from a cone of valid inequalities that is unbounded and, thus, needs to be truncated, or "normalized". We consider general normalizations defined by arbitrary closed convex sets and derive dual problems for generating L&P cuts. This unified theoretical framework generalizes and covers a wide group of already known normalizations. We also give conditions for proving finite convergence of the cutting-plane procedure that results from using such general L&P cuts.

Keywords: Lift-and-Project, disjunctive cuts, Branch-and-Cut, cutting planes, 0-1 integer programming.

AMS Subject classification: 90C11, 90C57,90C25

1. Introduction and Motivation.

Consider the following mixed integer linear programming problem:

$$\begin{cases} \min & c^\top x \\ & Ax \geq b \\ & x \geq 0 \\ & x_j \in \{0,1\}\,, j = 1, \ldots, p\,, \end{cases} \tag{1}$$

* Research supported by CNPq under Grant No.142835/97-8.

** Research supported by FAPERJ (Brazil) under Grant No.E26/150.581/00 and by CNPq (Brazil) under Grant No. 301800/96.

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $c, x \in \mathbb{R}^n$. The combinatorial nature of the integer variables can make it impossible to solve (1) by a direct method. Instead, one solves iteratively relaxations of (1). For example, the continuous (or linear) relaxation consists in replacing the integrality constraints $x_j \in \{0,1\}$ by $x_j \in [0,1]$. To obtain integer solutions, it is necessary to "tighten" relaxed problems along iterations, for instance by using Branch-and-Cut algorithms, [13]. This methodology combines a Branch-and-Bound strategy with the generation of cutting planes in order to work with progressively smaller feasible sets.

In this context, it is crucial to have an efficient implementation of the method used to generate cuts. However, one of the first cutting-planes methods, Gomory's algorithm [11], is known to have poor rates of convergence for general programs. In contrast, Lift-and-Project (L&P) methods, [2], [4], [5], [17], seem well adapted for problems without any particular structure. In fact, even though the apparently simple idea of using disjunctions was already considered in [3], the first efficient implementable techniques to obtain disjunctive cuts were given only recently, in [4]. The basic idea in [4] is to search for cuts in a dual cone of *valid inequalities*, associated to the current L&P relaxation (cf. Proposition 2.4 below). To ensure the well-definedness of the cut generation problem, the (unbounded) cone has to be intersected with a certain *normalization* set.

Although successfully implemented, the good behaviour of L&P methods is not completely understood yet from the theoretical point of view. More precisely, there are no clear criteria on how to choose the "best" normalization set.

This issue was first addressed in [10,9,18], where specific normalization sets are analyzed individually from the dual point of view. In this paper, we go further in the dual analysis by considering a large class of normalization sets, namely the class of closed convex sets. Our approach allows us to give a unified framework that includes most already known normalizations, as well as lays the foundation for introducing new normalization sets. We stress that in this paper we address purely theoretical questions, which are a prerequisite for the development of efficient implementations. We note that the first important step toward implementations is to show that a cutting-plane procedure using general L&P cuts has finite convergence. Conditions which ensure this property are given in Theorem 5.4 below.

The rest of the paper is organized as follows. We proceed with Sections 1.1-1.2, where we recall some concepts and results of Convex Analysis and cutting planes methods, and introduce our notation. The material on L&P methods, given in Section 2, allows us to specify L&P problems used to generate cuts. We present some results on duality in Section 3, and we state a general dual L&P cut generation problem, that we particularize to different choices of normalization sets in Section 4. Finally, in Section 5 we analyze the convergence for the resulting L&P cutting-planes algorithms. We also exhibit in Example 5.8 a case of convex normalization sets for which the cut generation problem is not well-defined.

1.1. Notation and basic results of Convex Analysis

To simplify the notation, we sometimes use abbreviated expressions, the meaning of which is explained below.

- For algebraic purposes we consider vectors to be column vectors. However, for the sake of simplicity, we sometimes use a compact (transposed) notation when referring to composite vectors. For instance, (x, y) stands for the column vector $\begin{pmatrix} x \\ y \end{pmatrix}$.
- When clear from the context, we use the short form $\{x = x_0\}$ to denote the set $\{(x, y) : x = x_0\} \subseteq \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$.

The rest of our notation is fairly standard in Convex Analysis. In particular, it closely follows [12].
Associated to a convex lower semicontinuous function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$,

- the *conjugate* of f is $f^*(\xi) := \sup_{z\in\mathbb{R}^n}\{\xi^\top z - f(z)\}$;
- the *effective domain* of f is $\operatorname{dom} f := \{x \in \mathbb{R}^n : f(x) < +\infty\}$;
- the *subdifferential* of f at a given point x is

$$\partial f(x) := \{g \in \mathbb{R}^n : f(y) \geq f(x) + g^\top(y - x) \text{ for all } y \in \mathbb{R}^n\}.$$

For an arbitrary set $\mathcal{X}$, and a convex set C in $\mathbb{R}^n$,

- the *shadow* projection of $\mathcal{X} \subset \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$, cf. [12, Example III.1.2.5]. is defined by

$$\operatorname{Proj}_{x_1} \mathcal{X} := \{x_1 : (x_1, x_2) \in \mathcal{X}\}.$$

  In other words, the shadow projection is the orthogonal projection onto the subspace $\mathbb{R}^{n_1}$.
- The *convex hull* of $\mathcal{X}$ is denoted by $\operatorname{conv} \mathcal{X}$.
- The *relative interior* of C is denoted by $\operatorname{ri} \mathcal{C}$. **ojo Before, you had it defined for $\mathcal{X}$ which is not convex**

– **ojo Before, this definition was formally incorrect** An *extreme point* of C is a point $x \in C$ which cannot be represented by a convex combination of points in C, all different from x itself. Specifically, there exist no integer l and $x \neq x_i \in C, \lambda_i \geq 0, i = 1, \ldots, l, \sum_{i=1}^{l} \lambda_i = 1$, such that $x = \sum_{i=1}^{l} \lambda_i x_i$. The set of all extreme points of C is denoted by $\operatorname{ext} C$.

– The *polar set* of C is $C^{\circ} := \{\xi \in \mathbb{R}^n : \xi^{\top} c \leq 1 \text{ for all } c \in C\}$. When C is a cone, the right-hand side in the inequality above can be set to zero.

In particular, for a polyhedral cone $\mathcal{C}$ defined by a homogenous system of linear inequalities with a matrix Z, we have:

$$\text{if } \mathcal{C} := \{\zeta : Z\zeta \geq 0\} \quad \text{then} \quad \mathcal{C}^{\circ} = \{z = Z^{\top}\lambda : \lambda \geq 0\}. \tag{2}$$

– The *support* function of C is $\sigma_C(z) := \sup\{z^{\top} c : c \in C\}$.

Given $C_1 \subset \mathbb{R}^{n_1}$ and $C_2 \subset \mathbb{R}^{n_2}$, for the Cartesian product $C_1 \times C_2$ we have:

$$\sigma_{C_1 \times C_2}(z_1, z_2) = \sigma_{C_1}(z_1) + \sigma_{C_2}(z_2). \tag{3}$$

– The *indicator* function of C is $I_C(z) := \begin{cases} 0 & \text{if } z \in C \\ +\infty & \text{otherwise.} \end{cases}$

The support and the indicator functions are conjugate to each other: $I_C^* = \sigma_C$. If, in addition, C is closed, then

$$\sigma_C(\hat{z}) = \hat{g}^{\top}\hat{z} \Rightarrow \hat{g} \in \partial\sigma_C(\hat{z}) \Rightarrow \sigma_{C^{\circ}}(\hat{g}) \leq 1 \tag{4}$$

(the first implication is by the definition of σ_C, and the rest follows from the fact that $\hat{g} \in \partial\sigma_C(\hat{z})$ if and only if $\hat{z} \in \partial{\sigma_C}^*(\hat{g}) = \partial I_C(\hat{g})$, which in turn implies that $\hat{g} \in C$, a condition equivalent to $\sigma_{C^{\circ}}(\hat{g}) \leq 1$).

– The *gauge function* γ_C is $\gamma_C(z) := \inf\{\lambda > 0 : z \in \lambda C\}$.

Whenever $0 \in \operatorname{int} C$, the interior of the set C, it holds that $\sigma_C = \gamma_{C^{\circ}}$ and

$$\gamma_C(z) \leq 1 \iff z \in C. \tag{5}$$

– Finally, we recall a result on Fenchel duality:

Theorem 31.4 in [16] *Given a convex lower semicontinuous function $\mathcal{F}$ and a non-empty convex cone $\mathcal{C}$, the primal problem* $\inf\{\mathcal{F}(u) : u \in \mathcal{C}\}$ *has a Fenchel* *dual* $-\inf\{\mathcal{F}^*(-s) : s \in \mathcal{C}^\circ\}$, *with both optimal values being equal. If, in addition,* $\mathcal{C}$ *is polyhedral and*

$$\operatorname{ri}\operatorname{dom}\mathcal{F} \cap \mathcal{C} \neq \emptyset \quad \textit{and} \quad \operatorname{ri}\operatorname{dom}\mathcal{F}^* \cap \mathcal{C}^\circ \neq \emptyset\,, \tag{6}$$

then both infima are attained, yielding

$$\min\{\mathcal{F}(u) : u \in \mathcal{C}\} = \max\{\mathcal{F}^*(-s) : s \in \mathcal{C}^\circ\}\,. \square$$

We will refer to conditions in (6) as primal and dual constraint qualification conditions (CQC). They ensure the existence of solutions of the primal and dual problems.

1.2. Generating cutting planes

Consider a cutting-planes algorithm, and let x^k be the current iterate. Let a pair $(\alpha^k, \beta^k) \in \mathbb{R}^n \times \mathbb{R}$ define a valid inequality ${\alpha^k}^\top x \geq \beta^k$. This defines a cut that separates x^k from the feasible set, if it holds that

$${\alpha^k}^\top x^k < \beta^k \qquad \text{and} \qquad {\alpha^k}^\top x \geq \beta^k \text{ for all } x \text{ feasible in (1).}$$

In an abstract form, valid inequalities are generated by solving

$$\begin{cases} \max\limits_{\alpha,\beta} f_k(\alpha, \beta) \\ \qquad (\alpha, \beta) \in \Lambda^k\,, \end{cases} \tag{7}$$

where the (constraint) set, $\Lambda^k \subset \mathbb{R}^n \times \mathbb{R}$ contains pairs (α, β) defining valid inequalities at x^k. The objective function depends on the point to be cut off and quantifies "how far" the generated cut is from x^k. Usual measures are given by $f_k(\alpha, \beta) := \beta - \alpha^\top x^k$ or by the Euclidean distance from x^k to the hyperplane $\{x \in \mathbb{R}^n : \alpha^\top x = \beta\}$. For (7) to be well defined, a CQC such as $(\operatorname{ri}\operatorname{dom} f_k) \cap \Lambda^k \neq \emptyset$ is assumed.

Since cuts are generated at each iteration, it is desirable that Λ^k should yield a simple problem (7). For instance, a linear or quadratic program, or a convex program. For the particular case of L&P cuts, Λ^k is a (truncated) cone arising naturally from the development, see (12) and (13) below.

In the following sections we consider how to use the L&P approach to generate cuts. Later, in Section 5, we study variants of cutting-planes algorithms using L&P cuts that we prove to be convergent in a finite number of iterations.

2. Lift-and-Project methods

Balas, Ceria and Cornuéjols proposed in [4,5] the Lift-and-Project method for solving 0-1 mixed integer linear programs. The method is related to previous work on disjunctive programming, matrix-cuts and other hierarchical relaxations, [3,1,14,17]. It has been extended recently to convex 0-1 programs, [19,10,18].

Essentially, the L&P method generates the convex hull of the feasible set in (1) by forcing at each iteration one of the integer components, say x_j, to satisfy the disjunction "either $x_j = 0$ or $x_j = 1$". By repeating this procedure enough number of times, the first p variables of x are forced to be binary after a finite number of iterations.

More precisely, given a polyhedral relaxation K of the feasible set in (1), the L&P method proceeds in two steps:

- "Lift" step: Multiply some of the constraints defining K by valid inequalities. This yields a nonlinear formulation for K that can be linearized by introducing new variables. The result is a polyhedral representation of the feasible set in a space of dimension higher than that of K.
- "Project" step: project the polyhedron obtained in the "Lift" step onto the original space.

The resulting polyhedron defines a relaxation of the feasible set in (1) that is contained in K (i.e., the relaxation is "tighter" than K). The same procedure, applied iteratively, yields the convex hull of the feasible set in (1) after a finite number of iterations, see [4, Corollary 2.3].

To better illustrate the technique, consider the following simple example. Let $K \subset \mathbb{R}^2$ be given by the system

$$K \quad \longleftrightarrow \quad \begin{cases} 1 + 2x_1 - 2x_2 \geq 0 \\ 3 - 2x_1 - 2x_2 \geq 0 \\ \qquad\quad x_1\,, x_2 \geq 0 \\ 1 - x_1\,, 1 - x_2 \geq 0\,. \end{cases}$$

There are $n := 2$ variables and $m := 6$ constraints. We now apply one L&P step to compute $P_1(K)$.

"Lift" step: multiply the inequalities above by $x_1 \geq 0$ and $1 - x_1 \geq 0$ to obtain the system

$$\begin{cases} 2x_1^2 - 2x_1x_2 + x_1 \geq 0 \\ -2x_1^2 - 2x_1x_2 + 3x_1 \geq 0 \\ x_1^2\,,\, x_1x_2 \geq 0 \\ -x_1^2 + x_1\,,\, -2x_1x_2 + x_1 \geq 0 \\ 1 - 2x_1^2 + 2x_1x_2 + x_1 - 2x_2 \geq 0 \\ 3 + 2x_1^2 + 2x_1x_2 - 5x_1 - 2x_2 \geq 0 \\ -x_1^2x_1\,,\, -x_1x_2 + x_2 \geq 0 \\ 1 + x_1^2 - 2x_1\,,\, 1 + x_1x_2 - x_1 - x_2 \geq 0\,. \end{cases}$$

To linearize this system, substitute the squares x_1^2 by x_1 (this does not affect the validity of the inequalities for the 0-1 points that we are interested in), and substitute the products x_1x_2 by a new variable y_2. After eliminating redundant and vacuous constraints and reordering the inequalities, we obtain the system:

$$\begin{cases} 1 - x_1 - 2x_2 + 2y_2 \geq 0 \\ 3 - 3x_1 - 2x_2 + 2y_2 \geq 0 \\ 3x_1 - 2y_2 \geq 0 \\ x_1 - 2y_2 \geq 0 \\ 1 - x_1 - x_2 + y_2 \geq 0 \\ x_2 - y_2 \geq 0 \\ 1 - x_1 \geq 0 \\ y_2 \geq 0\,. \end{cases} \tag{8}$$

This new system has $2n - 1 = 3$ variables and, although in this case we have eliminated many redundant inequalities in the intermediate "multiplying" step, it has $2m = 12$ inequalities.

"Project" step: $P_1(K)$ is obtained by projecting the polyhedron defined by (8) onto the (x_1, x_2)-space:

$$P_1(K) \quad \longleftrightarrow \quad \begin{cases} x_1 \geq 0 \\ x_2 \geq 0 \\ 1 - x_1 \geq 0 \\ 1 - 2x_2 \geq 0 \end{cases} \quad \text{for} \quad K \quad \longleftrightarrow \quad \begin{cases} 1 + 2x_1 - 2x_2 \geq 0 \\ 3 - 2x_1 - 2x_2 \geq 0 \\ x_1\,,\, x_2 \geq 0 \\ 1 - x_1\,,\, 1 - x_2 \geq 0\,. \end{cases}$$

A similar idea is underlying in [14] and [17], although the lifting steps therein involve spaces with higher dimension than the one used in [4].

For example, the method in [14] uses $n(n+1)2$ variables and $2nm$ constraints. All the methods in [4,14,17] can be considered of L&P type. For our study, we focus on the particular method of [4] and refer to it as *the* L&P method.

A potentially serious drawback of the methodology above is the projection step, whose implementation may be too expensive or even computationally impossible to accomplish. Yet, the obtained tighter relaxation is polyhedral and can be used to generate cuts which are useful within a Branch-and-Cut algorithm.

Next, we review some properties of these relaxations.

2.1. *The relaxations $P_j(K)$*

Given an initial polyhedron K such that $K \cap \mathbb{R}^p \subseteq [0,1]^p \times \mathbb{R}^{n-p}$ and a binary index $j \leq p$, consider the disjunctive polyhedron

$$P_j(K) := \operatorname{conv}\left((K \cap \{x_j \leq 0\}) \cup (K \cap \{x_j \geq 1\})\right) . \tag{9}$$

This relaxed polyhedron can be obtained by applying one step of the L&P procedure, see [4, Theorem 2.1]. Specifically, by using in the "Lift" step above the valid inequalities $x_j \geq 0$ and $1 - x_j \geq 0$.

As shown below, see (13), $P_j(K)$ yields well defined problems in (7). This is a consequence of the following result.

Proposition 2.1. Let $x \in \operatorname{ext} K$ be such that $0 < x_j < 1$. Then $x \notin P_j(K)$.

Proof. For contradiction purposes, suppose $x \in P_j(K)$. Since $P_j(K) \subseteq K$, this means that $x \in \operatorname{ext} P_j(K)$. The result follows from (9), because

$$x \in \operatorname{ext} P_j(K) = \operatorname{ext}\{x \in K : x_j = 0\} \cup \operatorname{ext}\{x \in K : x_j = 1\} ,$$

yielding that x_j is either 0 or 1. $\square$

It is possible to consider different representations of $P_j(K)$. For instance, $P_j(K)$ can be described by the feasible set of a certain linear programming problem, or even by means of certain dual cone $P_j^*(K)$ defined in (12) below.

We start with the linear programming representation, related to [3], characterizing $P_j(K)$ as the shadow projection of a higher-dimensional set.

Suppose the polyhedron K under consideration has the explicit form

$$K = \{x \in \mathbb{R}^n : \tilde{A}x \geq \tilde{b}\} ,$$

where $\tilde{A} \in \mathbb{R}^{\tilde{m} \times n}$ and $\tilde{b} \in \mathbb{R}^{\tilde{m}}$. To embed $P_j(K)$ in a higher-dimensional setting, we first introduce additional variables $y := (y^0, y^1) \in \mathbb{R}^n \times \mathbb{R}^n$ and $y_0 := (y_0^0, y_0^1) \in \mathbb{R} \times \mathbb{R}$ to represent it in $\mathbb{R}^{3n+2}$, following eq. (5) in [4]. Then, we add nonpositive perturbations $\varepsilon = (\varepsilon^0, \varepsilon^1) \in \mathbb{R}^{\tilde{m}} \times \mathbb{R}^{\tilde{m}}$ and $\varepsilon_0 = (\varepsilon_0^0, \varepsilon_0^1) \in \mathbb{R} \times \mathbb{R}$ to the right hand side terms and introduce a variable, $x_0 \in \mathbb{R}$, in order to allow $y_0^0 + y_0^1$ not to be restricted to 1. Altogether, consider the linear system on variables $(x, y, x_0, y_0, \varepsilon, \varepsilon_0) \in \mathbb{R}^{3n+3} \times \mathbb{R}^{2\tilde{m}} \times \mathbb{R}^2 = \mathbb{R}^{3n+2\tilde{m}+5}$ defined as follows, where e_j is the j-th unit vector in $\mathbb{R}^n$:

$$\begin{cases} y^0 + y^1 = x, \\ \tilde{A}y^0 - \tilde{b}y_0^0 \geq \varepsilon^0, \\ -e_j^\top y^0 \geq \varepsilon_0^0, \\ \tilde{A}y^1 - \tilde{b}y_0^1 \geq \varepsilon^1, \\ e_j^\top y^1 - y_0^1 \geq \varepsilon_0^1, \\ y_0^0 + y_0^1 = x_0, \\ y^0, y^1, y_0^0, y_0^1 \geq 0\,, \\ \varepsilon, \varepsilon_0 \leq 0\,. \end{cases} \tag{10}$$

Let $\Pi_j(K)$ denote the set of vectors $(x, x_0, \varepsilon, \varepsilon_0) \in \mathbb{R}^{n+2\tilde{m}+3}$ for which there exists a vector $(y, y_0) \in \mathbb{R}^{2n+2}$ such that $(x, x_0, y, y_0, \varepsilon, \varepsilon_0)$ is a solution of (10). Given $\varepsilon, \varepsilon_0$ fixed, we define $\Pi_j(K; \varepsilon, \varepsilon_0) := \{(x, x_0) : (x, x_0, \varepsilon, \varepsilon_0) \in \Pi_j(K)\}$. This set is a polyhedron in $\mathbb{R}^{n+1}$ that can be used to describe $P_j(K)$ using the shadow projection on the x's components:

Proposition 2.2. With the notation above,

$$P_j(K) = \mathrm{Proj}_x(\Pi_j(K; 0, 0) \cap \{x_0 = 1\})\,.$$

Proof. The result follows from [3, Theorem 2.1, Corollary 2.1.1], after the following change of notation:

notation in [3]	Q	A^0	A^1	a_0^0	a_0^1	ξ^h	ξ_0^h
ours	$\{0,1\}$	$\begin{bmatrix} \tilde{A} \\ -e_j \end{bmatrix}$	$\begin{bmatrix} \tilde{A} \\ e_j \end{bmatrix}$	$\begin{bmatrix} \tilde{b} \\ 0 \end{bmatrix}$	$\begin{bmatrix} \tilde{b} \\ 1 \end{bmatrix}$	y^h	y_0^h .

$\square$

Remark 2.3. When $\tilde{A} := A$ and $\tilde{b} := b$, the set K is the linear relaxation of (1). In this case $P_j(K)$ is a relaxation of the feasible set of (1). Actually, since for any nonpositive $\varepsilon, \varepsilon_0$ it holds that $\mathrm{Proj}_x(\Pi_j(K; \varepsilon, \varepsilon_0) \cap \{x_0 = 1\}) \supseteq P_j(K)$, the set $\mathrm{Proj}_x(\Pi_j(K) \cap \{x_0 = 1\})$ also provides relaxation of the feasible set in (1). Note, however, that this relaxation is "less tight" than $P_j(K)$. Finally, note that the result does not hold if we allow $\varepsilon, \varepsilon_0$ to have some positive component.

We pass now to the dual description of $P_j(K)$, using the shadow projection of a certain dual set $P_j^*(K)$. Instead of primal variables x, we consider pairs (α, β) defining valid inequalities, and additional variables $\mu := (\mu^0, \mu^1) \in \mathbb{R}^{\tilde{m}} \times \mathbb{R}^{\tilde{m}}$ and $\mu_0 := (\mu_0^0, \mu_0^1) \in \mathbb{R} \times \mathbb{R}$. Consider now the linear system on variables $(\alpha, \beta, \mu, \mu_0) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^{2\tilde{m}} \times \mathbb{R}^2 = \mathbb{R}^{n+2\tilde{m}+3}$ defined as

$$\left\{\begin{array}{llllll}\alpha & -\tilde{A}^\top\mu^0 + e_j\mu_0^0 & & & \geq 0, \\ \alpha & & -\tilde{A}^\top\mu^1 - e_j\mu_0^1 & & \geq 0, \\ \quad -\beta & +\tilde{b}^\top\mu^0 & & & \geq 0, \\ \quad -\beta & & +\tilde{b}^\top\mu^1 & +\mu_0^1 & \geq 0, \\ & & \mu^0,\ \mu^1,\ \mu_0^0,\ \mu_0^1 & & \geq 0\,. \end{array}\right. \tag{11}$$

The nonempty solution set of the system above, denoted by $\Pi_j^*(K)$, is a cone in $\mathbb{R}^{n+2\tilde{m}+3}$. Note that this system is related to the one in eq. (7) in [4, Thm. 2.10], with the notation (u, v, u_0, v_0) therein corresponding to $(\mu^0, \mu^1, \mu_0^0, \mu_0^1)$. The shadow projection of this cone (on the (α, β)'s components) defines the following set[1]

$$P_j^*(K) := \mathrm{Proj}_{(\alpha,\beta)}\left(\Pi_j^*(K)\right), \tag{12}$$

that allows a dual representation of $P_j(K)$ given below.

Proposition 2.4. [4, Theorem 2.10] With the notation above,

$$P_j(K) = \left\{x \in \mathbb{R}^n : \alpha^\top x \geq \beta \ \text{ for all } (\alpha, \beta) \in P_j^*(K)\right\}. \ \square$$

This result suggests the use of $P_j^*(K)$ in the constraint set Λ^k of (7). Cuts generated with this feasible set are called *L&P cuts*. To generate the cuts, we solve problem of the form

$$\left\{\begin{array}{l}\max\limits_{\alpha,\beta} f_k(\alpha, \beta) \\ \qquad (\alpha, \beta) \in \Lambda^k \subseteq P_j^*(K^k)\,, \end{array}\right. \tag{13}$$

where K^k is the current relaxation in the cutting-planes algorithm.

[1] Note that, even though $P_j^*(K)$ is called *(reverse) polar cone of* $P_j(K)$ in [3,1], it is *not* the standard polar set of Convex Analysis, which is $P_j(K)^\circ$, as defined in Section 1.1.

ojo: As a matter of English what follows is OK, as well as what is in the footnote. But the meaning is not the same... If Λ^k is the whole cone $P_j^*(K^k)$, Proposition 2.1 guarantees that (13) has a solution, i.e., a L&P cut exists. Since x^k is not feasible for (1), there exists some index j such that $x_j \neq 0, 1$. However, in the typical case when f_k is given by the affine function $\beta - \alpha^\top x^k$, defining Λ^k as the whole cone $P_j^*(K^k)$ would yield unbounded problems (13). In order to have well defined problems, it is preferable to consider smaller sets Λ^k. [2] The cone has to be truncated by adding *normalization* constraints, i.e., by letting $\Lambda^k := P_j^*(K^k) \cap S$ for a suitable chosen set S. We study such normalizations in a general setting and discuss some examples in the next section. Other choices for the measure f_k, including the Euclidean distance, are discussed in [15].

3. Some duality theory

Consider the normalized affine variant of (13) given by

$$\begin{cases} \max\limits_{\alpha,\beta} \beta - \alpha^\top x^k \\ \qquad (\alpha, \beta) \in P_j^*(K^k) \cap S\,. \end{cases} \tag{14}$$

In order to keep our discussion general, we only require the following (fairly weak) conditions on the normalization set.

Assumption 3.1. The normalization set $S \subset \mathbb{R}^n \times \mathbb{R}$ in (14) is nonempty, closed and convex. Furthermore, S is such that $P_j^*(K^k) \cap S$, the feasible set in (14), is bounded. □

If the set S is a polyhedron, (14) is a linear programming problem, whose size can be reduced without impairing the quality of the cut [4,9].

In addition, in [15] it is discussed how to generate "deep" cuts, obtained by replacing the affine objective function in (14) by a measure of the distance between the cutting plane and x^k, not necessarily the Euclidean distance. In this

[2] era: Note that when Λ^k is the whole cone $P_j^*(K^k)$, since x^k is not feasible for (1), there exists some index j such that $x_j \neq 0, 1$. Hence, by Proposition 2.1, (13) has a solution, i.e., a L&P cut exists. However, in order to have well defined problems, it is better to consider smaller sets Λ^k. Typically, when f_k is given by the affine function $\beta - \alpha^\top x^k$, defining Λ^k as the whole cone $P_j^*(K^k)$ would yield unbounded problems (13).

case, one is naturally led to consider convex (and not just polyhedral) normalization sets S.

The analysis of (14) gives rise to some well known normalization sets, to be considered in Sections 4.1, 4.2 below. Further choices are possible, if not only (α,β) but also other variables from (11) are considered (such is the case for the multipliers normalization given in Section 4.3 below).

We make use of (11) to write (14) with variables $(\alpha,\beta,\mu,\mu_0)\in\mathbb{R}^{n+2\tilde{m}+3}$, which yields

$$\begin{cases}\max\limits_{\alpha,\beta,\mu,\mu_0} \beta-\alpha^\top x^k \\ \qquad (\alpha,\beta,\mu,\mu_0)\in\Pi_j^*(K^k)\cap\Sigma\,.\end{cases} \tag{15}$$

The normalization set $\Sigma\subseteq\mathbb{R}^{n+2\tilde{m}+3}$ satisfies conditions analogous to those established in Assumption 3.1 for S. Note that (14) is a special case of (15), with normalization set $\Sigma:=S\times\mathbb{R}^{2\tilde{m}+2}$.

To find the dual of (15), we first prove a rather general result on duality.

We start with applying the Fenchel duality to the general problem (P) below, defined like (15), but with a general objective function F and no normalization set.

Theorem 3.2. Let $F:\mathbb{R}^{n+2\tilde{m}+3}\to\mathbb{R}\cup\{-\infty\}$ be a closed concave function satisfying the primal-dual CQC

$$\operatorname{ri}\operatorname{dom}F\cap\Pi_j^*(K^k)\neq\emptyset \quad\text{and}\quad \operatorname{ri}\operatorname{dom}(-F)^*\cap\Pi_j^*(K^k)^\circ\neq\emptyset\,.$$

The problem

$$\begin{cases}\max\limits_{\alpha,\beta,\mu,\mu_0} F(\alpha,\beta,\mu,\mu_0) \\ \qquad (\alpha,\beta,\mu,\mu_0)\in\Pi_j^*(K^k)\,.\end{cases} \tag{P}$$

has a Fenchel dual of the form

$$\begin{cases}\min\limits_{x,x_0,\varepsilon,\varepsilon_0} \Big(-F\Big)^*(x,-x_0,-\varepsilon,-\varepsilon_0) \\ \qquad (x,x_0,\varepsilon,\varepsilon_0)\in\Pi_j(K^k)\,.\end{cases} \tag{DP}$$

Proof. We apply the result on Fenchel duality stated at the end of Section 1.1, with $\mathcal{F}=-F$ and $\mathcal{C}=\Pi_j^*(K^k)$, a polyhedral cone. Our CQC correspond to those in (6). Hence, the optimal values are attained and the Fenchel dual of (P) is

$$\begin{cases} \min_{s} \Big(-F\Big)^{*}(-s) \\ \qquad s \in \Pi_j^*(K^k)^{\circ}\,. \end{cases}$$

To put this dual problem in the format of (DP), first split s into components, i.e., $s := (-x, x_0, \varepsilon, \varepsilon_0)$. Then use the result in (2). More precisely, $\mathcal{C}$ therein is $\Pi_j^*(K^k)$ and Z is the matrix in (11). Then apply (2) with $\zeta = (\alpha, \beta, \mu, \mu_0)$, $z = s$ and $\lambda = (y, y_0)$ to see that $s = (-x, x_0, \varepsilon, \varepsilon_0) \in \Pi_j^*(K^k)^{\circ}$ is equivalent to $(x, x_0, \varepsilon, \varepsilon_0) \in \Pi_j(K^k)$. □

Now we derive the dual for the L&P cut generation problem (15).

Corollary 3.3. Consider problem (15) and suppose that Σ is nonempty, closed and convex. Suppose further that the following CQC holds:

there exists a point $(x, x_0, \varepsilon, \varepsilon_0) \in \Pi_j(K^k)$ such that
$\sigma_\Sigma(x - x^k, 1 - x_0, -\varepsilon, -\varepsilon_0) < +\infty\,.$

Then it follows that (15) has the following Fenchel dual:

$$\begin{cases} \min_{x, x_0, \varepsilon, \varepsilon_0} \sigma_\Sigma(x - x^k, 1 - x_0, -\varepsilon, -\varepsilon_0) \\ \qquad (x, x_0, \varepsilon, \varepsilon_0) \in \Pi_j(K^k)\,. \end{cases} \tag{16}$$

Proof. Write (15) using the indicator function of Σ in the objective:

$$\begin{cases} \max_{\alpha, \beta, \mu, \mu_0} \beta - \alpha^\top x^k - I_\Sigma(\alpha, \beta, \mu, \mu_0) \\ \qquad (\alpha, \beta, \mu, \mu_0) \in \Pi_j^*(K^k)\,. \end{cases}$$

In order to apply Theorem 3.2, we show that both the primal and dual CQC hold for $F(\alpha, \beta, \mu, \mu_0) := \beta - \alpha^\top x^k - I_\Sigma(\alpha, \beta, \mu, \mu_0)$ and the conjugate $(-F)^*$. The primal CQC is always satisfied, since it corresponds to $\Sigma \cap \Pi_j^*(K^k) \neq \emptyset$.
To compute the conjugate function, we write $-F$ as a sum, $-F = \phi_1 + \phi_2$, where

$$\phi_1(\alpha, \beta, \mu, \mu_0) := \alpha^\top x^k - \beta = (\alpha, \beta, \mu, \mu_0)^\top (x^k, -1, 0, 0)$$

and $\phi_2(\alpha, \beta, \mu, \mu_0) := I_\Sigma(\alpha, \beta, \mu, \mu_0)$. From [12, Theorem X.2.3.1], the conjugate of a sum is the *infimal convolution* of ϕ_1^* and ϕ_2^*:

$$\Big(\phi_1 + \phi_2\Big)^*(z) = \inf_{\xi}\{\phi_1^*(\xi) + \phi_2^*(z - \xi)\}\,.$$

Using the definition of conjugate, we obtain

$$\phi_1^* = I_{(x^k,-1,0,0)} \quad \text{and} \quad \phi_2^* = \sigma_\Sigma \,,$$

respectively. Thus,

$$\begin{aligned}\Big(-F\Big)^*(x,-x_0,-\varepsilon,-\varepsilon_0) &= \inf_\xi \Big\{ I_{(x^k,-1,0,0)}(\xi) + \sigma_\Sigma\Big((x,-x_0,-\varepsilon,-\varepsilon_0)-\xi\Big)\Big\} \\ &= \sigma_\Sigma(x-x^k, 1-x_0, -\varepsilon, -\varepsilon_0)\,.\end{aligned}$$

Using the definition of $(-F)^*$ and $\Pi_j^*(K^k)^\circ$, it can be shown that the dual CQC in Theorem 3.2 corresponds to our CQC, and the result follows. □

Because (14) is a particular case of (15) where multipliers μ, μ_0 are unconstrained, the following result, a straightforward application of Corollary 3.3 with $\Sigma := S \times \mathbb{R}^{2\tilde{m}+2}$, shows that using the unconstrained multipliers corresponds to setting both ε and ε_0 to zero in the dual.

Corollary 3.4. Let $S \in \mathbb{R}^{n+1}$ be a nonempty closed convex set. Suppose further that the following CQC holds:

there exists $(x, x_0) \in \Pi_j(K^k; 0, 0)$ such that $\sigma_S(x - x^k, 1 - x_0) \leq +\infty$.

Then, the dual program of (14) is

$$\begin{cases} \min\limits_{x,x_0} \sigma_S(x-x^k, 1-x_0) \\ \qquad (x,x_0) \in \Pi_j(K^k;0,0)\,. \end{cases} \tag{17}$$

□

4. Normalization sets

We give here some examples of application of Theorem 3.2 and its corollaries, covering the so-called *natural* cases in [4], [9]:

- α is kept bounded,
- β has a fixed sign,
- multipliers μ e μ_0 satisfy some "normalized" constraint.

For all the examples, the CQC needed to find the Fenchel duals correspond to assuming that the normalization sets are chosen so that the cut generation problem is well defined and has a solution.

4.1. Normalization in α

The normalization obtained by requiring α to belong to the unit ball with a given norm is considered in [4,9,18]. A similar analysis can be carried out by taking $\alpha \in C \subseteq \mathbb{R}^n$, an arbitrary compact convex set such that $0 \in \operatorname{int} C$, and setting

$$S := C \times \mathbb{R}\,.$$

This choice of S, together with the result (5) on gauge functions, reduces (14) to

$$\begin{cases} \max\limits_{\alpha,\beta} \beta - \alpha^\top x^k \\ \qquad (\alpha,\beta) \in P_j^*(K^k), \\ \qquad \gamma_C(\alpha) \le 1\,, \end{cases}$$

We have that $\sigma_S = \sigma_C + \sigma_{\mathbb{R}}$. Since $\sigma_S = \gamma_{C^\circ}$ and $\sigma_{\mathbb{R}} = I_{\{0\}}$, $\sigma_C(x - x^k, 1 - x_0) = \gamma_{C^\circ}(x - x^k)$ and $x_0 \equiv 1$ in (17).

Altogether, from Corollary 3.4 and Proposition 2.2, we obtain the dual problem

$$\begin{cases} \min\limits_{x} \gamma_{C^\circ}(x - x^k) \\ \qquad x \in P_j(K^k)\,. \end{cases}$$

This choice of S generalizes the “LHS-Normalization” in [9, eqs. (5) and (6)], obtained for the particular case of C given by a unit ball.

4.2. Normalization given by a hyperplane

Given $(a, a_0) \in \mathbb{R}^n \times \mathbb{R}$ and a scalar $\hat\beta \neq 0$, define S in (14) as a hyperplane not passing through the origin:

$$S := \{(\alpha,\beta) : a^\top\alpha + a_0\beta = \hat\beta\}\,.$$

That is, in (14) we have

$$\begin{cases} \max\limits_{\alpha,\beta} \beta - \alpha^\top x^k \\ \qquad (\alpha,\beta) \in P_j^*(K^k), \\ \qquad a^\top\alpha + a_0\beta = \hat\beta\,. \end{cases} \tag{18}$$

We split our analysis by considering the cases $a_0 \neq 0$ and $a_0 = 0$.

• If $a_0 \neq 0$, then $s := (0, \frac{\hat{\beta}}{a_0}) \in S$ and

$$S = (0, \frac{\hat{\beta}}{a_0}) + \{(\alpha, \beta) : a^\top \alpha + a_0 \beta = 0\} =: s + \mathcal{H}\,.$$

Thus, $\sigma_S = \sigma_{\{s\}} + \sigma_{\mathcal{H}}$ and using [12, Example V.3.4.5], we obtain

$$\sigma_S(x - x^k, 1 - x_0) = \begin{cases} \frac{\hat{\beta}}{a_0}(1 - x_0) & \text{if } x - x^k = \frac{1 - x_0}{a_0}\, a\,, \\ +\infty & \text{otherwise.} \end{cases}$$

Corollary 3.4 gives the dual for (18):

$$\begin{cases} \min\limits_x \frac{\hat{\beta}}{a_0}(1 - x_0) \\ \quad (x, x_0) \in \Pi_j(K^k; 0, 0) \\ \quad x = x^k + \frac{1 - x_0}{a_0}\, a\,. \end{cases}$$

By choosing $a = 0$ and $a_0 = 1$, we recover the RHS-Normalization of [9].

• If $a_0 = 0$ then $a \neq 0$, $s := (\hat{\beta}/\|a\|^2\; a, 0) \in S$.

Define now $\mathcal{H} := \{(\alpha, \beta) : a^\top \alpha + a_0 \beta = 0\}$ to obtain

$$\begin{aligned} \sigma_S(x - x^k, 1 - x_0) &= \frac{\hat{\beta}}{\|a\|^2} \sigma_{\{(a,0)\}}(x - x^k, 1 - x_0) + \sigma_{\mathcal{H}}(x - x^k, 1 - x_0) \\ &= \begin{cases} \frac{\hat{\beta}}{\|a\|^2} a^\top (x - x^k) & \text{if } (x - x^k, 1 - x_0) = \lambda(a, 0)\,, \text{ for some } \lambda \in \mathbb{R} \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

As a result, $x_0 \equiv 1$ in (17). Therefore Corollary 3.4, Proposition 2.2 and (3), give the dual problem

$$\begin{cases} \min \lambda \hat{\beta} \\ \quad x := x^k + \lambda a \in P_j(K^k)\,. \end{cases}$$

By choosing $a \neq 0$, $\hat{\beta} := 1$, and $a_0 := 0$, we recover the normalization proposed in [6, Section 7].

4.3. Multipliers normalization

We study now normalization sets involving constraints on multipliers, forcing vectors (μ, μ_0) to belong to a general hyperplane. To this aim, given $0 \neq \bar{d} = (d, d_0) \in \mathbb{R}^{2\tilde{m}+2}$, define

$$\Sigma := \{(\alpha, \beta, \mu, \mu_0) : d^\top \mu + d_0^\top \mu_0 = 1\}\,.$$

Note that with this choice of Σ, (15) becomes

$$\begin{cases} \max\limits_{\alpha,\beta,\mu,\mu_0} \beta - \alpha^\top x^k \\ \qquad (\alpha, \beta, \mu, \mu_0) \in \Pi_j^*(K^k)\,, \\ \qquad d_\mu^\top \mu + d_0^\top \mu_0 = 1\,. \end{cases}$$

To compute the support function of Σ, let $\omega \in \mathbb{R}^{2\tilde{m}+2}$ be such that $\bar{d}^\top \omega = 1$, with $\bar{d} = (d, d_0)$, and write

$$\Sigma = \mathbb{R}^{n+1} \times \Big(\omega + \{(\mu, \mu_0) : d^\top \mu + d_0^\top \mu_0 = 0\}\Big)\,.$$

As a result, the objective function in (16) is

$$\sigma_\Sigma(x - x^k, 1 - x_0, -\varepsilon, -\varepsilon_0) = \begin{cases} \lambda & \text{if } x = x^k\,,\ x_0 = 1\,,\ \text{and } (\varepsilon, \varepsilon_0) = -\lambda \bar{d}\,, \\ +\infty & \text{otherwise,} \end{cases}$$

or equivalently,

$$\sigma_\Sigma(x - x^k, 1 - x_0, -\varepsilon, -\varepsilon_0) = \begin{cases} \lambda & \text{if } (x^k, 1) \in \Pi_j(K^k; -\lambda \bar{d})\,, \\ +\infty & \text{otherwise.} \end{cases} \tag{19}$$

Applying Corollary 3.3 and using (19), we obtain the dual problem

$$\begin{cases} \min\limits_{\lambda} \lambda \\ \qquad (x^k, 1) \in \Pi_j(K^k; -\lambda d, -\lambda d_0)\,. \end{cases}$$

Two of the normalizations in [9] are particular instances of this dual problem:

- Multipliers normalization: $\bar{d} := e$, the vector with all components equal to 1, and $\omega := \dfrac{1}{2m+2} e$.
- (u_0, v_0) normalization: $d := 0$, $d_0 := (1, 1)$, $\omega := (0, 0, \frac{1}{2}, \frac{1}{2})$.

To conclude, we give general conditions for having finite convergence of a L&P cutting-planes algorithm using general normalization sets.

5. On the convergence of L&P cutting-planes algorithms

Finite convergence for special L&P based cutting planes algorithms is already known when the normalization set $S \subseteq \mathbb{R}^n$ is polyhedral, [4, Theorem 3.1], [8, Theorem 5.1], or if the normalization is given by bounding the norm of α, [18]. Next we analyze the finite convergence of more general L&P cutting-planes algorithms. We consider a family of algorithms similar to the ones introduced in [4, p. 309] and [18, p. 74] (related also to the "Cut" algorithm from [7]) and analyze broader families of normalizations which enjoy properties that guarantee the convergence in a finite number of steps.

Before stating the algorithm pattern, we need some notation from [4]:

- A cut is called a *j-cut* if it is generated by solving a cut generation problem constructed from the disjunction "either $x_j = 0$ or $x_j = 1$".
- We let K_j^k denote the polyhedron defined by the union of $\{x : Ax \geq b\,, x \geq 0\}$ and all the i-cuts with $i \leq j$, generated up to iteration k. Note that, in particular, $K_0^k = K^1$ for all k.

L&P Cutting Plane algorithm pattern

Initialization. Let K^1 be the feasible set of the continuous relaxation of (1). Set $k = 1$.

LP resolution. Find x^k as an extreme (basic) solution of $\min\{c^\top x : x \in K^k\}$. If $x_j^k \in \{0, 1\}$ for all $j = 1, \ldots, p$, STOP.

Cut generation. Otherwise x^k is fractional. Let $j \in \{1, \ldots, p\}$ be the largest index such that $0 < x_j^k < 1$. Find a suitable j-cut ${\alpha^k}^\top x \geq \beta^k$.

Update. ojo: something is missing after x in formula for K $K^{k+1} := K^k \cap \{x :\}, k := k + 1$, loop to LP resolution. □

In the following, we assume that it is always possible to generate a j-cut, see Assumption 5.2 below. Depending on the normalization used, the j-cut is defined by solving a cut generation problem (CGP), either like (15) or like (16). In particular, the algorithms in [4] and [18] correspond to the following CGPs:

(i) The algorithm in [4] finds j-cuts by solving

$$\begin{cases} \max\limits_{\alpha,\beta} \beta - \alpha^\top x^k \\ \qquad (\alpha, \beta) \in P_j^*(K_{j-1}^k) \\ \qquad \|\alpha\| \leq 1\,, \end{cases}$$

where $\|\cdot\| = \|\cdot\|_1$ or $\|\cdot\| = \|\cdot\|_\infty$. This problem is just (14) with K^k replaced by K^k_{j-1} and $S := \{(\alpha,\beta) : \|\alpha\| \le 1\}$.

(ii) The algorithm in [18] finds j-cuts by solving

$$\begin{cases} \min\limits_{x} \|x - x^k\| \\ \qquad x \in P_j(K^k_{j-1})\,, \end{cases}$$

where $\|\cdot\|$ is any norm in $\mathbb{R}^n$. This problem is just (17) with K^k replaced by K^k_{j-1} and $S := \{(\alpha,\beta) : \|\alpha\| \le 1\}$ (this means in particular that $x_0 \equiv 1$).

For proving convergence we consider the following cases, generalizing (i) and (ii):

Polyhedral case. Let $\Sigma \subseteq \mathbb{R}^{n+2\tilde{m}+3}$ be a polyhedral set and consider the CGP: Find a basic solution $(\alpha^k, \beta^k, \mu^k, \mu^k_0)$ of

$$\begin{cases} \max\limits_{\alpha,\beta,\mu,\mu_0} \beta - \alpha^\top x^k \\ \qquad (\alpha,\beta,\mu,\mu_0) \in \Pi^*_j(K^k_{j-1}) \cap \Sigma\,. \end{cases} \tag{20}$$

Except for the requirement of finding a *basic* solution, (20) is simply (15) with K^k therein replaced by K^k_{j-1}.

Convex case. Let Σ' be a compact convex set in the (α,μ,μ_0)-space such that $0 \in \operatorname{int}\Sigma'$. Define $\Sigma := \Sigma' \times \mathbb{R}$ and consider the following CGP:

$$\begin{cases} \min\limits_{x,\varepsilon,\varepsilon_0} H(x,\varepsilon,\varepsilon_0) := \sigma_\Sigma(x - x^k, 0, -\varepsilon, -\varepsilon_0) \\ \qquad (x, 1, \varepsilon, \varepsilon_0) \in \Pi_j(K^k_{j-1})\,. \end{cases} \tag{21}$$

Let $(\hat{x}, \hat{\varepsilon}, \hat{\varepsilon}_0)$ be a solution of (21), with $(\alpha^k, 0, \mu^k, \mu^k_0) \in \Sigma$ realizing the optimal value:

$$H(\hat{x},\hat{\varepsilon},\hat{\varepsilon}_0) = \sigma_\Sigma(\hat{x} - x^k, 0, -\hat{\varepsilon}, -\hat{\varepsilon}_0) = (\alpha^k, 0, -\mu^k, -\mu^k_0)^\top(\hat{x} - x^k, 0, \hat{\varepsilon}, \hat{\varepsilon}_0)\,.$$

Finally, define $\beta^k := {\alpha^k}^\top \hat{x}$. □

We will show in Theorem 5.4 that a cutting-planes algorithm using normalization sets as in the Polyhedral and Convex Cases above converges to a solution of (1) in a finite number of steps.

Remark 5.1. Before proving this result, a few comments are in order. The reason to choose the "disjunctive" variable x_j by using a *lexicographic* order in the Cut generation step of the algorithm is used in the context of an inductive argument as follows. For the initial step, namely for $t = 1$, each 1-cut is related to a face of a *fixed* polyhedron. We show that each face of this polyhedron is generated a finite number of times. Then, there can only be a finite number of 1-cuts.

For an arbitrary $t > 1$, we make the inductive assumption that there are finitely many j-cuts for $j = 1, \ldots, t-1$, and again show that there is a polyhedron that remains fixed between the generation of two j-cuts, with $j < t$. Since the t-cuts generated are again related to faces of such fixed polyhedron, the fact that each face is generated only a finite number times yields, as before, that there are only finitely many t-cuts. Altogether, the key arguments in the reasoning are:

$$\begin{array}{ll} (i) & \text{cuts are related to faces of "fixed" polyhedra;} \\ (ii) & \text{each face of such polyhedra is generated a finite number of times.} \end{array} \tag{22}$$

In the Polyhedral case, the requirement that the solution of the CGP is basic ensures satisfaction of (22)(ii), since a linear program like CGP has a finite number of basis.

For the Convex case, normalizations need not be polyhedral, and we can no longer rely on basic solutions. Nevertheless, each cut defines by construction a support inequality for the fixed polyhedron in (22)(i). To prove that each face is generated finitely many times, we find a uniform estimate that bounds away from zero the angle between two different t-cuts defining the same face. This is Lemma 5.5 below, where we make intensive use of the dual information.

To state the main result of this section we make the following assumption.

Assumption 5.2. The inequality ${\alpha^k}^\top x \geq \beta^k$ produced in the Cut generation step is a valid inequality for (1) and cuts off the point x^k. In other words, ${\alpha^k}^\top x \geq \beta^k$ for all x in the feasible set in (1) but ${\alpha^k}^\top x^k < \beta^k$.

Remark 5.3. It can be shown that Assumption 5.2 always holds for the normalization defined by bounding α (cf. § 4.1) and for the "multipliers" or (u_0, v_0) normalizations in [9] (cf. § 4.3). It also holds for normalizations in § 4.2 whenever the feasible set in (18) is nonempty. However, there exist normalization families as in the Convex case for which further conditions are needed to guarantee the generation of cuts, see Example 5.8.

Theorem 5.4. Let Σ be a normalization set as in the Polyhedral or Convex cases satisfying Assumption 5.2. For the Convex case, assume further that Σ is unconstrained in β, i.e., $\Sigma := \Sigma' \times \mathbb{R}$, where Σ' is a convex compact set in the (α, μ, μ_0)-space containing the origin in its interior.

Then, the corresponding L&P cutting planes algorithm converges to a solution of (1) in a finite number of steps.

The proof of this result for the case of polyhedral normalizations is very similar to the one in [4, Theorem 3.1]. The convex case follows the lines of [18], but we extend Lemma 3.1 therein by the technical Lemma 5.5 below, which is crucial for proving convergence.

Given a closed convex set $\Sigma \subseteq \mathbb{R}^N$, define the constant

$$\cos\theta_\Sigma := \min_{w_1 : \|w_1\|=1} \sigma_\Sigma(w_1) \min_{w_2 : \|w_2\|=1} \sigma_{\Sigma^\circ}(w_2).$$

If Σ is compact and $0 \in \operatorname{int}\Sigma$, then this constant is positive and is the smallest satisfying

$$\sigma_\Sigma(w_1)\sigma_{\Sigma^\circ}(w_2) \leq \cos\theta_\Sigma \|w_1\|\|w_2\| \quad \forall\, w_1, w_2 \in \mathbb{R}^N .$$

We define the angle between two vectors $\hat{g}$ and g in $\mathbb{R}^N$ by

$$[0, \pi] \ni \theta_{\hat{g},g} := \arccos \frac{\hat{g}^\top g}{\|\hat{g}\|\|g\|} .$$

Lemma 5.5. Let Σ be a given compact convex set such that $0 \in \operatorname{int}\Sigma$. Let $U \subsetneq V \subseteq \mathbb{R}^N$ be two convex closed sets and $\bar{z} \in V \backslash U$. Assume that $\hat{z}$ is a solution of

$$\begin{cases} \min\limits_z \sigma_\Sigma(z - \bar{z}) \\ \quad z \in U . \end{cases}$$

and let $\hat{g}$ be a point in Σ for which the optimal value is attained, i.e., $\sigma_\Sigma(\hat{z} - \bar{z}) = \hat{g}^\top(\hat{z} - \bar{z})$. Then, for any $g \in \mathbb{R}^N$ such that $\sigma_V(-g) = -g^\top \hat{z}$, the angle between $\hat{g}$ and g is bounded from below by the positive constant $\theta_{\hat{g},g} \geq \dfrac{\pi}{2} - \theta_\Sigma > 0$.

Proof. By assumption, $\sigma_\Sigma(\hat{z} - \bar{z}) = \hat{g}^\top(\hat{z} - \bar{z})$; note that such $\hat{g}$ always exists, because the problem has a solution $\hat{z}$ whose optimal value is finite. This implies $\sigma_{\Sigma^\circ}(\hat{g}) \leq 1$, by (4). Then, letting $\hat{\theta} := \theta_{\hat{g},\hat{z}-\bar{z}}$, it holds that

$$\sigma_\Sigma(\hat{z} - \bar{z})\sigma_{\Sigma^\circ}(\hat{g}) \leq \hat{g}^\top(\hat{z} - \bar{z}) = \cos\hat{\theta}\|\hat{g}\|\|\hat{z} - \bar{z}\| . \tag{23}$$

The definition of support function and our assumption on Σ imply that

$$\sigma_\Sigma(\hat z - \bar z)\sigma_{\Sigma^\circ}(\hat g) \geq \left(\min_{w:\|w\|=1} \sigma_\Sigma(w) \min_{w_2:\|w_2\|=1} \sigma_{\Sigma^\circ}(w_2)\right) \|\hat z - \bar z\|\|\hat g\|$$
$$= \cos\theta_\Sigma \|\hat z - \bar z\|\|\hat g\| \,.$$

Together with (23), this means that $\cos\hat\theta \geq \cos\theta_\Sigma$. Hence,

$$0 \leq \hat\theta \leq \theta_\Sigma < \frac{\pi}{2} \,.$$

Now, take $g \in \mathbb{R}^N$ such that $\sigma_V(-g) = -g^\top \hat z$, so that

$$-g^\top \hat z \geq -g^\top z \quad \text{for all } z \in V \,. \tag{24}$$

The vectors g and $\hat z - \bar z$ can be rewritten using $\hat g$ and vectors h_g , $h_z \perp \hat g$ as follows:

$$g = \cos\theta_{\hat g, g}\hat g + \sin\theta_{\hat g,g} h_g \quad \text{and} \quad \hat z - \bar z = \cos\hat\theta \hat g + \sin\hat\theta h_z \,.$$

With this decomposition, $\|g\| = \|\hat g\| + \|h_g\|$ and $\|\hat z - \bar z\| = \|\hat g\| + \|h_z\|$, so that

$$\frac{g^\top(\hat z - \bar z)}{\|g\|\|\hat z - \bar z\|} = \cos\theta_{\hat g,g}\cos\hat\theta + \sin\theta_{\hat g,g}\sin\hat\theta \frac{h_g^\top h_z}{\|h_g\|\|h_z\|} \,.$$

Since $\sin\theta_{\hat g,g}\sin\hat\theta \geq 0$, the Cauchy-Schwarz inequality yields

$$\frac{g^\top(\hat z - \bar z)}{\|g\|\|\hat z - \bar z\|} \geq \cos\theta_{\hat g,g}\cos\hat\theta - \sin\theta_{\hat g,g}\sin\hat\theta = cos(\theta_{\hat g,g} + \hat\theta) \,.$$

Using (24) with $z = \bar z$, the latter implies that $\cos(\theta_{\hat g,g} + \hat\theta) \leq 0$. Since $0 < \frac{\pi}{2} - \theta_\Sigma \leq \frac{\pi}{2} - \hat\theta \leq \theta_{\hat g,g}$, the result follows. $\square$

We shall use the following particular case of the preceding lemma.

Corollary 5.6. Let Σ' be a given compact convex set such that $0 \in \operatorname{int}\Sigma$. Let $\tilde U \subseteq \mathbb{R}^{n_1 \times n_2}$ be a convex closed set and $U' := \operatorname{Proj}_{z_1}\tilde U \subseteq \mathbb{R}^{n_1}$ its shadow projection on $\mathbb{R}^{n_1}$. Let $\bar z_1 \in \mathbb{R}^{n_1} \backslash U'$ and define $\bar z = (\bar z_1, 0) \in \mathbb{R}^{n_1 \times n_2}$. Assume that $\hat z' = (\hat z_1', \hat z_2')$ is a solution of

$$\begin{cases} \min\limits_z \sigma_{\Sigma'}(z - \bar z) \\ \qquad z \in \tilde U \,. \end{cases}$$

Let $\hat g' = (\hat g_1', -\hat g_2')$ be a point in Σ' for which the optimal value is attained, i.e., $\sigma_{\Sigma'}(\hat z' - \bar z) = \hat g'^\top(\hat z' - \bar z)$. Then, for any $g_1 \in \mathbb{R}^n$ such that $\sigma_{U'}(-g_1) = -g_1^\top \hat z_1'$ and $g_1^\top \hat z_1' \leq g_1^\top \bar z_1$, the angle between $\hat g_1'$ and g_1 is bounded from below by the positive constant $\dfrac{\pi}{2} - \theta_{\Sigma'} > 0$. $\square$

We now prove Theorem 5.4 following the reasoning sketched in Remark 5.1.

Proof of Theorem 5.4. We organize our proof in two parts.
Claim 1. *While the algorithm does not stop, the Cut Generation step is solvable and generates a cut.*
By construction, $x^k \in \operatorname{ext} K^k$. Due to the "lexicographical" choice of j, using the same argument as in the proof of [4, Theorem 3.1] and Proposition 2.1, it can be shown that $x^k \in \operatorname{ext} K_j^k$ but $x^k \notin P_j(K_{j-1}^k)$. By Assumption 5.2, CGP is well defined and Claim 1 follows.
Claim 2. *For each $t = 1, \ldots, p$, the algorithm generates a finite number of t-cuts.*
We prove this claim by induction on t. We choose a family of polyhedra satisfying (i) and (ii) in (22). The family depends on the normalization set involved:
Polyhedral case. The family of polyhedra is

$$\left\{W_{t-1}^k := \Pi_t^*(K_{t-1}^k) \cap \Sigma : t = 1, \ldots, p\,; k \geq 1\right\}.$$

Convex case. Given a polyhedron $K = \{x \in \mathbb{R}^n : Ax \geq b\}$, where A is an $m \times n$ matrix and $b \in \mathbb{R}^m$, let

$$U_t^k := \operatorname{Proj}_{(x,\varepsilon,\varepsilon_0)}\left(\Pi_t(K) \cap \{x_0 = 1\}\right).$$

The family of polyhedra is $\left\{U_t^k : t = 1, \ldots, p\,; k \geq 1\right\}$. Note that $(x^k, 0, 0) \notin U_t^k$. We now show, by induction, that both of those families satisfy (22).

When $t = 1$, first note that $K_0^k = K^1$, the initial relaxation for every k. Hence, W_1^k (or U_1^k depending on the case) remains fixed, because $W_1^k = \Pi_t^*(K_{t-1}^k) \cap \Sigma$ (or $U_1^k = \operatorname{Proj}_{(x,\varepsilon,\varepsilon_0)}\left(\Pi_1(K^1) \cap \{x_0 = 1\}\right)$). Then, for the polyhedral case, the feasible set in (20) is $W_1 := W_1^k$ for all k, whereas in the convex case the feasible set in (21) is $U_1 := U_1^k$ for all k.

In the polyhedral case, $W_1 := W_1^k$ for all k, so each 1-cut corresponds to a basis defining W_1. The result follows, since there are finitely many of such basis. In the convex case, $U_1 := U_1^k$ for all k, so each 1-cut corresponds to vector defining a face of U_1. Given a face of U_1, let $(\alpha^k, \beta^k := {\alpha^k}^\top \hat{x})$ and (α, β) be two different cuts corresponding to the face, and suppose that (α, β) was generated first. Apply Corollary 5.6 to

$$\tilde{U} = U_1,\ \bar{z}_1 = (x^k),\ \hat{z} = (\hat{x}, (\hat{\varepsilon}, \hat{\varepsilon}_0)),\ \hat{g} = (\alpha^k, (\mu^k, \mu_0^k)),\ g = (\alpha, (\mu, \mu_0)).$$

The assumption that (α, β) was generated first implies that
$\beta = \alpha^\top \hat{x}$ because (α, β) defines a face of of $\operatorname{Proj}_x U_1 = P_1(K^1)$ that is supporting at $\hat{x}$. This, in turn, implies that $\sigma_{U'}(-g_1) = -g_1^\top \hat{z}_1'$, and

$g_1^\top \hat{z}_1' \le g_1^\top \bar{z}_1$, because $x^k \in K^k$ and then, $\alpha^\top \hat{x} \le \alpha^\top x$.
Therefore, the chosen g satisfies the assumptions of Corollary 5.6 and we can conclude that $\theta_{\alpha^k,\alpha} \ge \frac{\pi}{2} - \theta_{\Sigma'}$. Since U_1 is a polyhedron, conclusion for the case $t = 1$ follows.

Make the inductive assumption that Claim 2 holds for $j = 1, \ldots, t-1$. Then, for all $j \in \{1, \ldots, t-1\}$, the set K_{t-1}^k is modified in the Update step of the algorithm by adding, at most a finite number of times, some j-cuts with $j < t$. We now show that between any such redefinitions, at most a finite number of t-cuts are generated. Between these redefinitions, the set K_{t-1}^k does not change; so, both sets W_t^k and U_t^k remain the same.

In the polyhedral case, the result follows by the reasoning similar to the case $t = 1$, showing that W_t^k remains fixed and that each t-cut corresponds to a basis defining W_t^k.

In the convex case, since U_t^k remains the same, each t-cut defines a face of this set. We can apply Corollary 5.6 as for the case $t = 1$, but for $\tilde{U} = U_t^k$. The result is that between modifications, each face of U_t^k is generated at most a finite number of times. Since U_t^k is a polyhedron, Claim 2 follows. □

Remark 5.7. It is shown in [7], for the case of Fenchel cuts, that bigger values of the corresponding θ_Σ lead to faster convergence. Since we generate L&P-cuts using relaxations $P_j(K^k)$ that change during the enumeration process, a similar conclusion may not hold in our setting. In any case, the notion of "big value" of θ_Σ can be misleading, because it depends on the norm used. For instance, consider the norm $\|\alpha\|_Q^2 = \langle \alpha, \alpha \rangle_Q := \alpha^\top Q \alpha$ induced by a positive definite matrix Q, and take $\Sigma := \{\alpha : \|\alpha\|_Q \le 1\}$. Then

$$\theta_\Sigma = \arccos\left(\frac{\lambda_{\min}(Q)}{\lambda_{\max}(Q)}\right),$$

where $\lambda_{\min}(Q)$ and $\lambda_{\max}(Q)$ are the minimal and maximal eigenvalues of Q, respectively. The use of the scalar product $\langle \cdot, \cdot \rangle_Q$ in the proof of Theorem 5.4 yields

$$\cos\theta_\Sigma := \min_{w:\|w\|_Q=1} \sigma_\Sigma(w) \min_{w_2:\|w_2\|_{Q^{-1}}=1} \sigma_{\Sigma^\circ}(w_2) = 0,$$

so $\theta_\Sigma = \pi/2$.

Recall that in the proof above, θ_Σ is used only to limit the number of different inequalities defining the same face of $P_j(K^k)$. We believe that a more significant bound could be

$$\cos\theta_\Sigma := \min_Q \left(\min_{w:\|w\|_Q=1} \sigma_\Sigma(w) \min_{w_2:\|w_2\|_{Q^{-1}}=1} \sigma_{\Sigma^\circ}(w_2) \right) ,$$

where the minimum is taken over the set of positive definite matrices Q.

We finish with an example showing that for normalization sets as in the **Convex case**, the CGP (21) may not generate a cut (i.e., Assumption 5.2 does not hold).

Example 5.8. Let $K^k \subset \mathbb{R}^2$ be the polyhedron defined by the system

$$\begin{cases} x_1 + x_2 \geq 1 \\ -2x_1 + 4x_2 \geq 1 \\ x_1\,,\ x_2 \geq 0 \\ x_1\,,\ x_2 \leq 1 \end{cases}$$

Consider a convex normalization set $\Sigma = \Sigma' \times \mathbb{R}$, with Σ' in the (α, μ, μ_0)-space to be defined below. We have that $\alpha\,, \mu^0 \in \mathbb{R}^2$ and, since the polyhedron is defined by $\tilde{m} := 6$ constraints, $\mu = (\mu^0, \mu^1) \in \mathbb{R}^6 \times \mathbb{R}^6$. Given a positive parameter M, let $(\alpha, \mu^0, \mu^1, \mu_0) \in \Sigma'$ be elements in the following ellipsoid:

$$M\alpha^\top\alpha + M{\mu^0}^\top\mu^0 + {\mu^1}^\top \begin{pmatrix} M &&&&& \\ & 1 &&&& \\ && 1 &&& \\ &&& M && \\ &&&& M & \\ &&&&& 1 \end{pmatrix} \mu^1 + {\mu_0}^\top \begin{pmatrix} M & \\ & 1 \end{pmatrix} \mu_0 \leq 1\,,$$

where in the matrices above we only specify nonzero elements.

Let $x^k := (0.5\,,\ 0.5)^\top$ be the basic solution which we want to cut off by solving the CGP (21), written for $j = 1$. Let $(\hat{x}, \hat{\varepsilon}, \hat{\varepsilon}_0)$ denote a solution of this problem. For each $(x, 1, \varepsilon, \varepsilon_0) \in \Pi_1(K^k)$ the objective support function is the maximum over Σ' of the scalar product

$$\alpha^\top(x - x^k) + 0 - \mu^\top\varepsilon - \mu_0^\top\varepsilon_0\,. \tag{25}$$

By choosing M big enough, the terms that will prevail in (25) are those corresponding to components having a factor M in the ellipsoid:

$$\alpha \qquad \mu^0 \qquad \mu^1_1 \qquad \mu^1_4 \qquad \mu^1_5 \qquad \mu^0_0 \,.$$

Therefore, the minimization in (21) will force the corresponding components of the optimal solution to be zero:

$$\hat{x} - x^k = 0 \qquad \hat{\varepsilon}^0 = 0 \qquad \hat{\varepsilon}^1_1 = 0 \qquad \hat{\varepsilon}^1_4 = 0 \qquad \hat{\varepsilon}^1_5 = 0 \qquad \hat{\varepsilon}^0_0 = 0 \,.$$

Altogether, (14) has a solution

$$\hat{x} = x^k \,, \quad \hat{\varepsilon}^0 = 0 \,, \quad \hat{\varepsilon}^1 = (0, -0.15, 0, 0, 0, 0) \,, \quad \hat{\varepsilon}_0 = (0, -0.45) \,,$$

which, together with $\hat{y} = (0, 0.05, 0.5, 0.45)$, $\hat{y}_0 = (0.05, 0.95)$, $\hat{x}_0 = 1$, solves (10). Since $\hat{x} = x^k$, the inequality $\alpha^\top x \geq \alpha^\top \hat{x}$ never cuts off x^k, and this is true for any value of α.

6. Concluding Remarks

We have presented a duality analysis for the L&P cut generation problems and an analysis of general conditions for the convergence of L&P cutting-planes algorithms to a solution of (1) in a finite number of iterations. The Polyhedral and Convex cases considered to prove finite convergence are related by duality, and it is interesting to note that our proof uses primal arguments for the Polyhedral case, and dual arguments for the Convex case.

We have considered general normalization sets and extended previous theoretical analysis of Ceria and Soares. Our approach covers all the normalizations already studied in the literature. Since the duality analysis is fairly general, it also opens the way to define L&P cuts not previously considered.

Furthermore, as pointed out by one of the referees, the dual theory in Section 3 can be extended to general disjunctive sets (not just with the form (9)).

Finally, we have given an example to show that even when CQC holds, a compact convex normalization set may not yield a well defined cut generation problem. This subject deserves further investigation. We are currently working in the effective implementation of new L&P cuts, to assess our approach with computational experiments.

Acknowledgments We are thankful to the referees, whose careful reading and constructive remarks helped us in improving the paper.

References

[1] E. Balas. Disjunctive programming. *Annals of Discrete Mathematics*, 5:3–51, 1979.

[2] E. Balas. A modified lift-and-project procedure. *Mathematical Programming*, 79:19–31, 1997.

[3] E. Balas. Disjunctive programming: Properties of the convex hull of feasible points. *Discrete Applied Mathematics*, 89:3–44, 1998. Originally, MSRR No. 348, Carnegie Mellon University, 1974.

[4] E. Balas, S. Ceria, and G. Cornuéjols. A lift-and-project cutting plane algorithm for mixed 0-1 programs. *Mathematical Programming*, 58:295–324, 1993.

[5] E. Balas, S. Ceria, and G. Cornuéjols. Mixed 0-1 programming by lift-and-project in a branch-and-cut framework. *Management Science*, 42:1229–1246, 1996.

[6] E. Balas and M. Perregaard. Lift-and-project for mixed 0-1 programming: Recent progress. Management Science Research Report MSRR-627, Carnegie Mellon University, Pittsburgh, PA, September 1999.

[7] E. A. Boyd. On the convergence of fenchel cutting planes in mixed-integer programming. *SIAM Journal on Optimization*, 5:421–435, 1995.

[8] S. Ceria. *Lift-and-Project methods for mixed 0-1 programs.* PhD thesis, Carnegie Mellon University, 1993.

[9] S. Ceria and J. Soares. Disjunctive cut generation for mixed 0-1 programs: duality and lifting. Columbia University, March 1997.

[10] S. Ceria and J. Soares. Convex programming for disjunctive programming. *Mathematical Programming*, 86:595–614, 1999.

[11] R. Gomory. Outline of an algorithm for integer solutions of linear programs. *Bulletin of the American Mathematical Society*, 64:275–278, 1958.

[12] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms.* Number 305-306 in Grund. der math. Wiss. Springer-Verlag, 1993. (two volumes).

[13] M. Jünger, G. Reinelt, and S. Thienel. Practical problem solving with cutting plane algorithms in combinatorial optimization. In W. Cook, L. Lovász, and P. Seymour, editors, *Combinatorial Optimization*, volume 20 of *DIMACS series in discrete mathematics and theoretical computer science*, pages 111–154. 1995.

[14] L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1:166–190, 1991.

[15] P.A. Rey. *Análise convexa e métodos lift-and-project para programação 0-1 (in portuguese).* Tese de doutorado, Depto. Engenharia Elétrica. PUC-Rio, April 2001.

[16] R.T. Rockafellar. *Convex Analysis.* Number 28 in Princeton Mathematics Ser. Princeton University Press, 1970.

[17] H. Sherali and W. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3:411–430, 1990.

[18] J. Soares. *Disjunctive convex programming.* PhD thesis, Columbia University, 1998.

[19] R. Stubbs and S. Mehrotra. A branch-and-cut method for 0-1 mixed convex programming. *Management Science*, 86:515–532, 1999.