



Refined cut selection for benders decomposition: applied to network capacity expansion problems

René Brandenberg¹  · Paul Stursberg¹

Received: 11 May 2020 / Revised: 10 July 2021 / Accepted: 1 October 2021 /
Published online: 13 November 2021
© The Author(s) 2021

Abstract

In this paper, we present a new perspective on cut generation in the context of Benders decomposition. The approach, which is based on the relation between the alternative polyhedron and the reverse polar set, helps us to improve established cut selection procedures for Benders cuts, like the one suggested by Fischetti et al. (Math Program Ser B 124(1–2):175–182, 2010). Our modified version of that criterion produces cuts which are always supporting and, unless in rare special cases, facet-defining. We discuss our approach in relation to the state of the art in cut generation for Benders decomposition. In particular, we refer to Pareto-optimality and facet-defining cuts and observe that each of these criteria can be matched to a particular subset of parametrizations for our cut generation framework. As a consequence, our framework covers the method to generate facet-defining cuts proposed by Conforti and Wolsey (Math Program Ser A 178:1–20, 2018) as a special case. We conclude the paper with a computational evaluation of the proposed cut selection method. For this, we use different instances of a capacity expansion problem for the European power system.

Keywords Benders decomposition · Decomposition methods · Cutting planes · Reverse polar set · Alternative polyhedron · Pareto optimal cuts · Facet defining cuts

1 Introduction

Consider a generic optimization problem with two subsets of variables x and y where x is restricted to lie in some set $S \subseteq \mathbb{R}^n$ and x and y are jointly constrained by a set

✉ René Brandenberg
rene.brandenberg@tum.de

Paul Stursberg
paul.stursberg@tum.de

¹ Technical University of Munich, Munich, Germany

of m linear inequalities. Such a problem can be written in the following form:

$$\begin{aligned}
 & \min c^\top x + d^\top y \\
 & \text{s.t. } Hx + Ay \leq b \\
 & \quad x \in S \\
 & \quad y \in \mathbb{R}^k
 \end{aligned} \tag{1.1}$$

The *interaction matrix* $H \in \mathbb{R}^{m \times n}$ captures the influence of the x -variables on the y -subproblem: For fixed $x^* \in \mathbb{R}^n$, (1.1) reduces to an ordinary linear program with constraints $Ay \leq b - Hx^*$, where $A \in \mathbb{R}^{m \times k}$, $y \in \mathbb{R}^k$, and $b - Hx^* \in \mathbb{R}^m$.

We are interested in cases where the size of the complete problem (1.1) leads to infeasibly high computation times (or memory demands), but both the problem over S and the problem resulting from fixing x can separately be solved much more efficiently due to their special structures. To deal with such problems, Benders (1962) introduced a method that works by iterating between these two “easier” problems:

For a problem of the form (1.1), let the function $z : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ represent the value of the optimal y -part of the objective function for a fixed vector x :

$$z(x) := \min_{y \in \mathbb{R}^k} \left\{ d^\top y \mid Ay \leq b - Hx \right\} \tag{1.2}$$

The corresponding epigraph of z is

$$\text{epi}(z) = \left\{ (x, \eta) \in \mathbb{R}^n \times \mathbb{R} \mid \exists y \in \mathbb{R}^k : \begin{array}{l} Ay \leq b - Hx \\ d^\top y \leq \eta \end{array} \right\}. \tag{1.3}$$

Writing $\text{epi}_S(z) := \text{epi}(z) \cap (S \times \mathbb{R})$, this provides us with an alternative representation of the optimization problem (1.1):

$$\min \left\{ c^\top x + \eta \mid (x, \eta) \in \text{epi}_S(z) \right\}$$

This representation suggests the following iterative algorithm: Start by finding a solution $(x^*, \eta^*) \in S \times \mathbb{R}$ that minimizes $c^\top x + \eta$ without any additional constraints (adding a generous lower bound for η to make the problem bounded). If $(x^*, \eta^*) \in \text{epi}(z)$, then $(x^*, \eta^*) \in \text{epi}_S(z)$ (since $x^* \in S$) and the solution is optimal. Otherwise, we add constraints violated by (x^*, η^*) but satisfied by all $(x', \eta') \in \text{epi}(z)$ and iterate. This is of course just an ordinary cutting plane algorithm and the crucial question is how to select a *separating inequality* in each iteration.

The original Benders algorithm uses *feasibility cuts* (cuts with coefficient 0 for the variable η) and *optimality cuts* (cuts with non-zero coefficient for the variable η), depending on whether or not the subproblem that results from fixing the x -variables is feasible (see, e. g., Vanderbeck and Wolsey 2010). Fischetti et al. (2010), on the other hand, present a unified perspective that covers both cases: They begin by observing

that the subproblem can be seen as a pure feasibility problem, represented by the set

$$\left\{ y \in \mathbb{R}^k \mid \begin{array}{l} Ay \leq b - Hx^* \\ d^\top y \leq \eta^* \end{array} \right\}. \tag{1.4}$$

This polyhedron will be empty if and only if $(x^*, \eta^*) \notin \text{epi}(z)$ and any Farkas certificate for emptiness of (1.4) can be used to derive an additional valid inequality. The set of such certificates (up to positive scaling)

$$P(x^*, \eta^*) := \left\{ (\gamma, \gamma_0) \geq 0 \mid \begin{array}{l} \gamma^\top A + \gamma_0 d^\top = 0 \\ \gamma^\top (b - Hx^*) + \gamma_0 \eta^* = -1 \end{array} \right\} \tag{1.5}$$

is called *alternative polyhedron*.¹ Thus $P(x^*, \eta^*) = \emptyset$ if and only if $(x^*, \eta^*) \in \text{epi}(z)$ and every point $(\gamma, \gamma_0) \in P(x^*, \eta^*)$ induces an inequality $\gamma^\top (b - Hx) + \gamma_0 \eta \geq 0$ that is valid for $\text{epi}(z)$ but violated by (x^*, η^*) .

This characterization is very useful and has been demonstrated empirically to work well in Fischetti et al. (2010). However, it exposes some fundamental issues, which are demonstrated by the following example (cf. Fig. 1).

Example 1.1 Consider the following optimization problem:

$$\begin{aligned} \min \quad & x + y && (1.6) \\ & -2x - y \leq -5 \\ & -\frac{1}{2}x - y \leq -3 \\ & -4x - 4y \leq -14 \\ & x \in S := \mathbb{R}_{\geq 0} \\ & y \in \mathbb{R} \end{aligned}$$

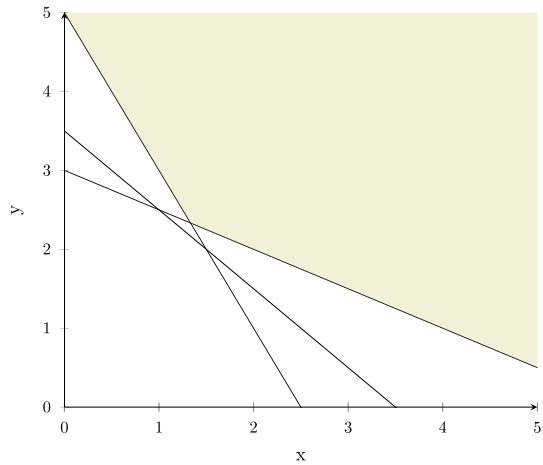
Note that the constraint $-4x - 4y \leq -14$ is redundant and does not support the feasible region. Suppose that we want to decompose the problem into its x -part and its y -part.

Writing the components of (γ, γ_0) in order $(\gamma_1, \gamma_2, \gamma_3, \gamma_0)$, we obtain

$$P(x^*, \eta^*) := \left\{ \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_0 \end{pmatrix} \geq 0 \mid \begin{array}{l} \gamma_0 - \gamma_1 - \gamma_2 - 4\gamma_3 = 0 \\ \gamma_0 \eta^* + \gamma_1(-5 + 2x^*) + \gamma_2(-3 + \frac{1}{2}x^*) + \gamma_3(-14 + 4x^*) = -1 \end{array} \right\}.$$

¹ Note that the specific form of the alternative polyhedron depends on the formulation of the basic optimization problem (1.1). In our case, a point $P(x^*, \eta^*)$ certifies infeasibility of (1.4) by the following observation: Let y be any vector contained in the set from (1.4). Then, for any $(\gamma, \gamma_0) \geq 0$ it holds that $\gamma^\top Ay + \gamma_0 d^\top y \leq \gamma^\top (b - Hx^*) + \gamma_0 \eta^*$. On the other hand, any $(\gamma, \gamma_0) \in P(x^*, \eta^*)$ implies $\gamma^\top Ay + \gamma_0 d^\top y = 0 > -1 = \gamma^\top (b - Hx^*) + \gamma_0 \eta^*$, independently of the choice of y .

Fig. 1 Constraints and feasible region for the optimization problem (1.6)



The three vertices of $P(0, 0)$ are

$$P_1 = \left(\frac{1}{5}, 0, 0, \frac{1}{5} \right)$$

$$P_2 = \left(0, \frac{1}{3}, 0, \frac{1}{3} \right)$$

$$P_3 = \left(0, 0, \frac{1}{14}, \frac{2}{7} \right).$$

As Gleeson and Ryan (1990) showed, each of these points corresponds to a minimal infeasible subsystem of (1.6) with the objective function written in inequality form $x + y \leq 0$. Consequently, each vertex yields one of the original inequalities as a cut. This notably includes the redundant inequality $-4x - 4y \leq -14$, which does not support the feasible region but is derived from the vertex P_3 in the alternative polyhedron (and furthermore minimizes the linear objective $\mathbb{1}$).

A cut generated from a point in the alternative polyhedron may thus be very weak, not even supporting the set $\text{epi}(z)$. This is true even if we use a vertex of the alternative polyhedron and even if that vertex minimizes a given linear objective such as the vector $\mathbb{1}$ as suggested in Fischetti et al. (2010).

In the following, we present an improved approach for cut generation in the context of Benders decomposition. Our method can be parametrized by the selection of an objective vector in primal space and produces facet cuts without any additional computational effort for all but a sub-dimensional set of parametrizations. In addition, our method is more robust with respect to the formulation of the problem than the original approach from Fischetti et al. (2010). In particular it always generates supporting cuts, avoiding the problem pointed out in the context of Example 1.1.

Our method is based on the relation between the alternative polyhedron as introduced above, which is commonly used in the context of Benders cut generation, and

the reverse polar set, originally introduced by Balas and Ivanescu (1964) in the context of transportation problems.

We show that the alternative polyhedron can be viewed as an extended formulation of the reverse polar set, providing us with a parametrizable method to generate cuts with different well-known desirable properties, most notably facet-defining cuts. As a special case, we obtain an (arguably simpler) alternative proof for the method to generate facet-defining cuts proposed by Conforti and Wolsey (2018), if applied to Benders decomposition. Our work links their approach more directly to previous work on cut selection, within the context of Benders decomposition (e.g., Fischetti et al. (2010)) as well as more generally for separation of convex sets (e.g., Cornuéjols and Lemaréchal (2006)).

Before we proceed by investigating different representations of the set of possible Benders cuts, it is useful to record a general characterization of the set of normal vectors for cuts separating a point from $\text{epi}(z)$ as defined in (1.3). In the following, we

say that a halfspace $H_{(\pi, \alpha)}^{\leq} := \left\{ x \in \mathbb{R}^n \mid \pi^\top x \leq \alpha \right\}$ is x -separating for a convex set $C \subset \mathbb{R}^n$ and a point $x \in \mathbb{R}^n \setminus C$ if $x \notin H_{(\pi, \alpha)}^{\leq} \supset C$.

Theorem 1.2 *Let z be defined as in (1.2) such that $\text{epi}(z) \neq \emptyset$ and let $(x^*, \eta^*), (\pi, \pi_0) \in \mathbb{R}^n \times \mathbb{R}$. Then (π, π_0) is the normal vector of a (x^*, η^*) -separating halfspace for $\text{epi}(z)$ if and only if there exists a vector $\gamma \in \mathbb{R}_{\geq 0}^m$ satisfying*

$$(\pi^\top, \pi_0) \begin{pmatrix} x^* \\ \eta^* \end{pmatrix} - \gamma^\top b > 0 \tag{1.7}$$

$$\gamma^\top A - \pi_0 d^\top = 0 \tag{1.8}$$

$$\gamma^\top H = \pi^\top \tag{1.9}$$

$$\pi_0 \leq 0. \tag{1.10}$$

Proof Let $h_{\text{epi}(z)}(\pi, \pi_0) := \sup \left\{ \pi^\top x + \pi_0 \eta \mid (x, \eta) \in \text{epi}(z) \right\}$ be the support function of $\text{epi}(z)$ evaluated at (π, π_0) . The vector (π, π_0) is the normal vector of a (x^*, η^*) -separating halfspace for $\text{epi}(z)$ if and only if

$$0 < (\pi^\top, \pi_0) \begin{pmatrix} x^* \\ \eta^* \end{pmatrix} - h_{\text{epi}(z)}(\pi, \pi_0). \tag{1.11}$$

By the definition of $\text{epi}(z)$ (which is closed and polyhedral) and then by strong LP duality, we obtain

$$h_{\text{epi}(z)}(\pi, \pi_0) = \max_{\substack{x \in \mathbb{R}^n, y \in \mathbb{R}^k \\ \eta \in \mathbb{R}}} \left\{ (\pi^\top, \pi_0) \begin{pmatrix} x \\ \eta \end{pmatrix} \mid \begin{array}{l} Ay \leq b - Hx \\ d^\top y \leq \eta \end{array} \right\} \tag{1.12}$$

$$= \min_{\substack{\gamma_0 \in \mathbb{R}_{>0} \\ \gamma \in \mathbb{R}_{\geq 0}^m}} \left\{ \gamma^\top b \mid \begin{array}{l} \gamma^\top A + \gamma_0 d^\top = 0 \\ \gamma^\top H = \pi^\top \\ -\gamma_0 = \pi_0 \end{array} \right\}. \tag{1.13}$$

Note that in order for the equality $-\gamma_0 = \pi_0$ to hold and (1.13) to be feasible (and hence (1.12) to be bounded), we need that $\pi_0 \leq 0$. Thus the optimality of any pair (γ, γ_0) for (1.13) is equivalent to the fulfillment of conditions (1.7) to (1.10). \square

As one can see from the proof above, any γ satisfying (1.8) to (1.10) is an upper bound for $h_{\text{epi}(z)}$. This means that given a certificate γ to prove that (π, π_0) is a normal vector of an (x^*, η^*) -separating halfspace $H_{((\pi, \pi_0), \alpha)}^\leq$, we immediately obtain a corresponding right hand side $\alpha := \gamma^\top b$. Furthermore, the definition of the support function $h_{\text{epi}(z)}$ immediately tells us when this right-hand side is actually optimal and the resulting halfspace supports $\text{epi}(z)$:

Remark 1.1 Let $(x^*, \eta^*) \in \mathbb{R}^n \times \mathbb{R}$ and let (π, π_0) be the normal vector of an (x^*, η^*) -separating halfspace for $\text{epi}(z)$. If γ minimizes $\gamma^\top b$ among all possible certificates in Theorem 1.2, then the halfspace $H_{((\pi, \pi_0), \gamma^\top b)}^\leq$ supports the set $\text{epi}(z)$.

2 Benders cuts from the reverse polar set

While it would be sufficient in the context of Benders decomposition to obtain an *arbitrary* (x^*, η^*) -separating halfspace whenever the set in (1.4) is empty, the alternative polyhedron $P(x^*, \eta^*)$ actually completely characterizes the set of *all* possible normal vectors of such halfspaces:

Corollary 2.1 *The alternative polyhedron (1.5) completely characterizes all normal vectors of (x^*, η^*) -separating halfspaces for $\text{epi}(z)$. In particular:*

- (a) *Let $(\gamma, \gamma_0) \in P(x^*, \eta^*)$. Then $\gamma^\top Hx - \gamma_0 \eta \leq \gamma^\top b$ is violated by (x^*, η^*) , but satisfied by all $(x, \eta) \in \text{epi}(z)$.*
- (b) *Let (π, π_0) be the normal vector of a (x^*, η^*) -separating halfspace for $\text{epi}(z)$. Then there exist $(\gamma, \gamma_0) \in P(x^*, \eta^*)$ and $\lambda \geq 0$ such that $(\gamma^\top H, -\gamma_0) = \lambda \cdot (\pi^\top, \pi_0)$.*

Observe, however, that in contrast to Remark 1.1, Corollary 2.1 does not guarantee that the cut generated from a point in the alternative polyhedron is supporting (as seen in Example 1.1 not even if that point is a vertex): A given vector $(\gamma, \gamma_0) \in P(x^*, \eta^*)$ might not minimize $\gamma^\top b$ among all points in $P(x^*, \eta^*)$ which lead to the same cut normal.

Alternatively, as argued by Cornuéjols and Lemaréchal (2006), the *reverse polar* of a convex set characterizes the set of normals of cuts that separate the origin from the set. The reverse polar was originally introduced in Balas and Ivanescu (1964) and can be defined as follows:

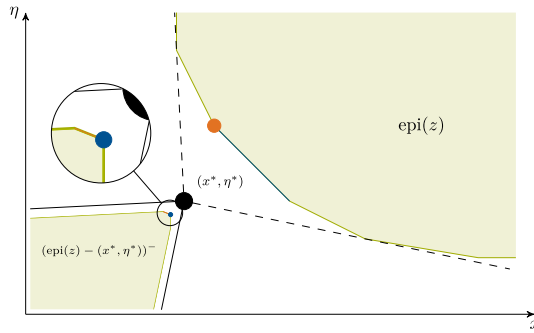


Fig. 2 The reverse polar set $(\text{epi}(z) - (x^*, \eta^*))^-$ and the corresponding polar cone (drawn in a coordinate system with (x^*, η^*) as the origin). It can be seen that $(\text{epi}(z) - (x^*, \eta^*))^-$ is contained in the polar cone $\text{pos}(\text{epi}(z) - (x^*, \eta^*))^\circ$ (indicated by the black solid lines) but offers a “richer” boundary from which we can choose cut normals. Specifically, for each vertex v of $(\text{epi}(z) - (x^*, \eta^*))^-$ there exists a facet of $\text{epi}(z)$ with normal vector v and vice versa (see Theorem 3.2)

Definition 2.1 Let $C \subseteq \mathbb{R}^n$ be a convex set. Then the *reverse polar set* C^- of C is defined as

$$C^- := \left\{ c \in \mathbb{R}^n \mid c^\top x \leq -1 \text{ for all } x \in C \right\}.$$

It is thus a subset of the *polar cone*

$$\text{pos}(C)^\circ := \left\{ c \in \mathbb{R}^n \mid c^\top x \leq 0 \text{ for all } x \in C \right\}.$$

We can use Theorem 1.2 and an appropriate positive scaling² to obtain the following description of the reverse polar set.³ Keeping in mind that we want to separate the point (x^*, η^*) rather than the origin, we must translate the set such that (x^*, η^*) becomes the origin (cf. Fig. 2).

$$(\text{epi}(z) - (x^*, \eta^*))^- = \left\{ (\pi, \pi_0) \in \mathbb{R}^n \times \mathbb{R} \mid \exists \gamma \geq 0 : \begin{aligned} &(\pi^\top, \pi_0) \begin{pmatrix} x^* \\ \eta^* \end{pmatrix} - \gamma^\top b \geq 1 \\ &A^\top \gamma - \pi_0 d = 0 \\ &H^\top \gamma = \pi \\ &\pi_0 \leq 0 \end{aligned} \right\} \tag{2.1}$$

Note furthermore that, as a consequence of Remark 1.1, we can compute for any given normal vector (π, π_0) a supporting inequality (if one exists) by solving problems (1.12) or (1.13).

² The set defined by the inequalities (1.7) to (1.10) is homogenous.

³ Since we use additional γ variables for the description, it is essentially an *extended formulation*.

We thus have at our disposal two alternative characterizations of the set of possible normal vectors of (x^*, η^*) -separating halfspaces: The alternative polyhedron and the reverse polar set. Despite their similarity, subtle differences exist between both representations that affect their usefulness for the generation of Benders cuts.

It should be noted at this point that we are not the first ones to notice the similarity between the approaches of Cornuéjols and Lemaréchal (2006) and Fischetti et al. (2010). Indeed, the work of Cornuéjols and Lemaréchal (2006) is explicitly cited in Fischetti et al. (2010), albeit only in a remark about the possibility to exchange normalization and objective function in optimization problems over the alternative polyhedron (see Corollary 2.4 below).

Before we proceed, we introduce a variant of the alternative polyhedron, the *relaxed alternative polyhedron*, which is also used in Gleeson and Ryan (1990). We will see that it is equivalent to the original alternative polyhedron for almost all purposes, but can more easily be connected to the reverse polar set:

Definition 2.2 Let a problem of the form (1.1) and a point $(x^*, \eta^*) \in \mathbb{R}^n \times \mathbb{R}$ be given. The *relaxed alternative polyhedron* $P^{\leq}(x^*, \eta^*)$ is defined as

$$P^{\leq}(x^*, \eta^*) := \left\{ (\gamma, \gamma_0) \geq 0 \mid \begin{array}{l} \gamma^\top A + \gamma_0 d^\top = 0 \\ \gamma^\top (b - Hx^*) + \gamma_0 \eta^* \leq -1 \end{array} \right\}.$$

To motivate the above definition, observe that optimization problems over the original and the relaxed alternative polyhedron are equivalent, provided that the optimization problem over the relaxed alternative polyhedron has a finite non-zero optimum:

Remark 2.1 Let z be defined as in (1.2) and let $(x^*, \eta^*) \in \mathbb{R}^n \times \mathbb{R}$. Let $(\tilde{\omega}, \tilde{\omega}_0) \in \mathbb{R}^m \times \mathbb{R}$ be such that $\max \left\{ \tilde{\omega}^\top \gamma + \tilde{\omega}_0 \gamma_0 \mid \gamma, \gamma_0 \in P^{\leq}(x^*, \eta^*) \right\} < 0$. Then the sets of optimal solutions for $\tilde{\omega}^\top \gamma + \tilde{\omega}_0 \gamma_0$ over $P^{\leq}(x^*, \eta^*)$ and $P(x^*, \eta^*)$ are identical. Furthermore, every vertex of $P^{\leq}(x^*, \eta^*)$ is also a vertex of $P(x^*, \eta^*)$.

The following key theorem now almost becomes a trivial observation. However, to our knowledge, the relation between the alternative polyhedron and the reverse polar set has not been made explicit in a similar fashion before (for a set S , we write

$$AS := \left\{ Ax \mid x \in S \right\}.$$

Theorem 2.1 Let z be defined as in (1.2) and $(x^*, \eta^*) \in \mathbb{R}^n \times \mathbb{R}$. Then

$$(\text{epi}(z) - (x^*, \eta^*))^- = \begin{pmatrix} H^\top & 0 \\ 0 & -1 \end{pmatrix} \cdot P^{\leq}(x^*, \eta^*).$$

Proof

$$\begin{pmatrix} H^\top & 0 \\ 0 & -1 \end{pmatrix} \cdot P^{\leq}(x^*, \eta^*) = \left\{ (H^\top \gamma, -\gamma_0) \mid \begin{array}{l} \gamma, \gamma_0 \geq 0 \\ \gamma^\top A + \gamma_0 d^\top = 0 \\ \gamma^\top (b - Hx^*) + \gamma_0 \eta^* \leq -1 \end{array} \right\}$$

$$\begin{aligned}
 &= \left\{ (\pi, \pi_0) \in \mathbb{R}^n \times \mathbb{R}_{\geq 0} \mid \exists \gamma \geq 0 : \begin{array}{l} H^\top \gamma = \pi \\ \gamma^\top A - \pi_0 d^\top = 0 \\ \gamma^\top b - \pi^\top x^* - \pi_0 \eta^* \leq -1 \end{array} \right\} \\
 &= (\text{epi}(z) - (x^*, \eta^*))^-
 \end{aligned}$$

□

One common scenario for Benders decomposition is where the *master* problem is significantly smaller than the *subproblem*. In this case Theorem 2.1 implies that the relaxed alternative polyhedron is an *extended formulation* for the reverse polar set, which in particular is always polynomial in size.

We revisit Example 1.1 to illustrate this observation.

Example 1.1 (continuing from p. 3) In the situation of the optimization problem (1.6), observe that the relaxed alternative polyhedron is

$$P^{\leq}(x^*, \eta^*) := \left\{ \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_0 \end{pmatrix} \geq 0 \mid \begin{array}{l} \gamma_0 - \gamma_1 - \gamma_2 - 4\gamma_3 = 0 \\ \gamma_0 \eta^* + \gamma_1(-5 + 2x^*) + \gamma_2(-3 + \frac{1}{2}x^*) + \gamma_3(-14 + 4x^*) \leq -1 \end{array} \right\}$$

It can be verified that $P^{\leq}(0, 0)$ is hence a 3-dimensional unbounded polyhedron in \mathbb{R}^4 with extremal rays $\text{pos}(0, 1, 0, 1)$, $\text{pos}(1, 0, 0, 1)$, $\text{pos}(0, 0, 1/4, 1)$ and with the original alternative polyhedron $P(0, 0)$ as the only bounded facet. We can now use Theorem 2.1 to derive the reverse polar set:

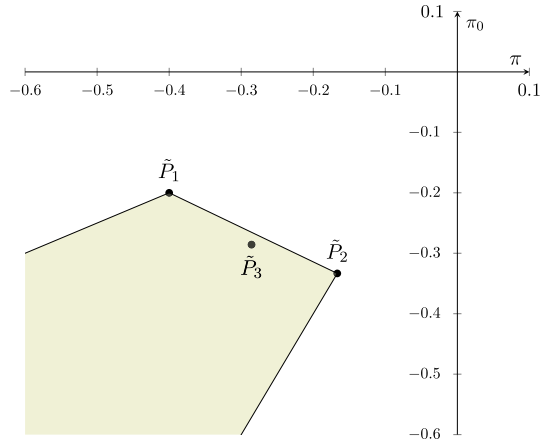
$$\begin{aligned}
 \text{epi}(z)^- &= (\text{epi}(z) - (0, 0))^- = \begin{pmatrix} H^\top & 0 \\ 0 & -1 \end{pmatrix} P^{\leq}(0, 0) \\
 &= \text{conv} \left(\left(-\frac{2}{5}, -\frac{1}{5} \right), \left(-\frac{1}{6}, -\frac{1}{3} \right), \left(-\frac{2}{7}, -\frac{2}{7} \right) \right) \\
 &\quad + \text{pos} \left(\left(-\frac{1}{2}, -1 \right), (-2, -1), (-1, -1) \right).
 \end{aligned}$$

The set $\text{epi}(z)^-$ is visualized in Fig. 3.

2.1 Cut-generating linear programs

One way to select a particular cut normal from the reverse polar set or the alternative polyhedron is by maximizing a linear objective function over these sets. Using Theorem 2.1, we can derive the precise relation between optimization problems over the reverse polar set and the alternative polyhedron.

Fig. 3 The set $\text{epi}(z)^-$ from Example 1.1. We can see that the point P_3 , which lead to the non-supporting cut above, is mapped to the interior of the reverse polar set and will hence not appear as an extremal solution



Corollary 2.2 Let z be defined as in (1.2), $(x^*, \eta^*), (\omega, \omega_0) \in \mathbb{R}^n \times \mathbb{R}$ and

$$(\tilde{\omega}, \tilde{\omega}_0) := (H\omega, -\omega_0). \tag{2.2}$$

Then (π, π_0) is an optimal solution to the problem

$$\max \left\{ \omega^\top \pi + \omega_0 \pi_0 \mid (\pi, \pi_0) \in (\text{epi}(z) - (x^*, \eta^*))^- \right\} \tag{2.3}$$

if and only if there exists γ such that $H^\top \gamma = \pi$ and $(\gamma, -\pi_0)$ is an optimal solution to the problem

$$\max \left\{ \tilde{\omega}^\top \gamma + \tilde{\omega}_0 \gamma_0 \mid (\gamma, \gamma_0) \in P^\leq(x^*, \eta^*) \right\}. \tag{2.4}$$

Furthermore, the objective values of both optimization problems are identical.

Proof Let (π, π_0) be an optimal solution to (2.3). By Theorem 2.1, there exists a vector γ with $H^\top \gamma = \pi$ such that $(\gamma, -\pi_0) \in P^\leq(x^*, \eta^*)$. Now, let (γ', γ'_0) be an arbitrary point in $P^\leq(x^*, \eta^*)$. By Theorem 2.1, $(H^\top \gamma', -\gamma'_0) \in (\text{epi}(z) - (x^*, \eta^*))^-$ and thus from the optimality of (π, π_0) for (2.3) we obtain

$$\begin{aligned} \tilde{\omega}^\top \gamma' + \tilde{\omega}_0 \gamma'_0 &= (H\omega)^\top \gamma' - \omega_0 \gamma'_0 = \omega^\top (H^\top \gamma') + \omega_0 (-\gamma'_0) \\ &\leq \omega^\top \pi + \omega_0 \pi_0 = (H\omega)^\top \gamma - \omega_0 (-\pi_0) = \tilde{\omega}^\top \gamma + \tilde{\omega}_0 (-\pi_0), \end{aligned}$$

which proves the optimality of $(\gamma, -\pi_0)$ for (2.4).

Similarly, let (γ, γ_0) be an optimal solution to (2.4), $\pi := H^\top \gamma$, and $\pi_0 := -\gamma_0$, which means by Theorem 2.1, $(\pi, \pi_0) \in (\text{epi}(z) - (x^*, \eta^*))^-$.

Now, let (π', π'_0) be an arbitrary point in $(\text{epi}(z) - (x^*, \eta^*))^-$. By Theorem 2.1, there exists γ' with $H^\top \gamma' = \pi'$ such that $(\gamma', -\pi'_0) \in P^\leq(x^*, \eta^*)$. Using the optimality of

(γ, γ_0) for (2.4), we obtain

$$\begin{aligned} \omega^\top \pi' + \omega_0 \pi'_0 &= (H\omega)^\top \gamma' - \omega_0(-\pi'_0) = \tilde{\omega}^\top \gamma' + \tilde{\omega}_0(-\pi'_0) \\ &\leq \tilde{\omega}^\top \gamma + \tilde{\omega}_0 \gamma_0 = \omega^\top H^\top \gamma - \omega_0 \gamma_0 = \omega^\top \pi + \omega_0 \pi_0, \end{aligned}$$

which proves the optimality of (π, π_0) for (2.3). □

Note that the optimization problem stated in (2.4) is technically more general, since there is no reason to limit ourselves to objective functions of the form (2.2) *a priori*. If we choose a different objective function, we still obtain a valid cut. However, since there may be no objective function (ω, ω_0) such that the resulting cut normal is optimal for (2.3), we lose some of the properties associated with optimal solutions from the reverse polar set.

Indeed, this is the approach that Fischetti et al. (2010) take: They use the problem in (2.4) with $\tilde{\omega}_m = 0$ for all m which correspond to rows of zeros in the interaction matrix H and $\tilde{\omega}_m = 1$ for all other m , as well as $\tilde{\omega}_0 = 1$ (or $\tilde{\omega}_0 = \kappa$ for some scaling factor $\kappa > 0$). In general, there exists no vector (ω, ω_0) such that this choice can be obtained by (2.2).

We now take a closer look at the role of objective functions in the context of Example 1.1:

Example 1.1 (continuing from p. 3) In the situation of the optimization problem (1.6), remember that the point P_3 actually minimizes the 1-norm over $P(0, 0)$ and is hence the unique result of the (unscaled) selection procedure from Fischetti et al. (2010). On the other hand the transformation from Theorem 2.1 actually maps this point, which lead to a non-supporting cut, to the interior of the reverse polar set. It will therefore never appear as an optimal solution of any linear optimization problem.

In order to obtain a supporting cut, we only have to make sure that the used objective can be written in the form $(H\omega, -\omega_0)$. In our example, if we choose the objective function over the alternative polyhedron from the set

$$\left\{ \left(\left(\begin{pmatrix} -2 \\ -1/2 \\ -4 \end{pmatrix} \cdot \omega, -\omega_0 \right) \mid \omega, \omega_0 \in \mathbb{R} \right) \right\},$$

then the point $P_3 \in P(0, 0)$ is never optimal.

To illustrate this further, we solve the optimization problem using $(\omega, \omega_0) := (1, 1)$ as an example, which implies that

$$(\tilde{\omega}, \tilde{\omega}_0) = (H\omega, -\omega_0)^\top = \left(\left(\begin{pmatrix} -2 \\ -1/2 \\ -4 \end{pmatrix} \cdot 1, -1 \right) \right).$$

Assuming that we begin with 0 as an initial lower bound for both x and η to make the problem bounded, we thus obtain $(x^1, \eta^1) := (0, 0)$ as an initial tentative solution.

The cut-generating problem is

$$\max \left\{ -2\gamma_1 - \frac{1}{2}\gamma_2 - 4\gamma_3 - \gamma_0 \mid (\gamma, \gamma_0) \in P(0, 0) \right\}$$

with optimal solution $P_1 = (1/5, 0, 0, 1/5)$. The resulting cut is $\gamma^\top Hx - \gamma_0 \eta \leq \gamma^\top b$, which resolves to

$$-\frac{2}{5}x - \frac{1}{5}y \leq -1,$$

the (linearly scaled) first inequality from (1.6).

Adding this inequality to the master problem, we obtain $(x^2, \eta^2) := (5/2, 0)$ as the next tentative solution. Now,

$$P(5/2, 0) = \left\{ \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_0 \end{pmatrix} \geq 0 \mid \begin{array}{l} \gamma_0 - \gamma_1 - \gamma_2 - 4\gamma_3 = 0 \\ -\frac{7}{4}\gamma_2 - 4\gamma_3 = -1 \end{array} \right\}$$

which leads to the cut-generating problem

$$\max \left\{ -2\gamma_1 - \frac{1}{2}\gamma_2 - 4\gamma_3 - \gamma_0 \mid (\gamma, \gamma_0) \in P(5/2, 0) \right\} = (0, 4/7, 0, 4/7)$$

from which we obtain the next cut:

$$-\frac{4}{14}x - \frac{4}{7}y \leq -\frac{12}{7},$$

the (linearly scaled) second inequality from (1.6).

In the next iteration, we obtain $(x^*, \eta^*) = (4/3, 7/3)$, the optimal solution. Correspondingly,

$$P(4/3, 7/3) = \left\{ \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_0 \end{pmatrix} \geq 0 \mid \begin{array}{l} \gamma_0 - \gamma_1 - \gamma_2 - 4\gamma_3 = 0 \\ \frac{7}{3}\gamma_0 - \frac{7}{3}\gamma_1 - \frac{7}{3}\gamma_2 - \frac{26}{3}\gamma_3 = -1 \end{array} \right\} = \emptyset,$$

which certifies optimality of the solution $(x, y) = (4/3, 7/3)$.

We have thus solved the optimization problem in two iterations, whereas the selection procedure from Fischetti et al. (2010) would have selected the point P_3 in the first iteration, leading to a cut that corresponds to the the redundant inequality in the original problem. It would thus require at least one additional iteration to solve the problem.

On the other hand, we will see that the fact that our approach yields two facet-defining cuts is not a coincidence: The following corollary shows that the generated

cuts are always at least supporting and we will see in Sect. 3.2 that the generated cuts are actually almost always facet-defining.

One interesting difference between the alternative polyhedron and the reverse polar set, which can be verified using the above example, is their different behavior with respect to algebraic operations on the set of inequalities: If, for instance, we scale one of the inequalities by a positive factor, the reverse polar set remains unchanged (just as the feasible region defined by the set of inequalities). On the other hand, the alternative polyhedron does change and, as a consequence, might yield a different optimal solution with respect to a given objective. In fact, by scaling the inequalities of (1.6) appropriately, we can actually avoid that the point P_3 is optimal for the approach from Fischetti et al. (2010). If an objective function is used which does not take this scaling into account, such as the vector of zeros and ones proposed by Fischetti et al. (2010), then the selected cut might change depending on the scaling factor. Even selecting a suitable manual scaling factor κ as mentioned above cannot fix this, since it cannot scale individual constraints against each other.

Combining our results from this section, we obtain the following statement:

Corollary 2.3 *Let z be defined as in (1.2) and (x^*, η^*) , $(\omega, \omega_0) \in \mathbb{R}^n \times \mathbb{R}$, $(\tilde{\omega}, \tilde{\omega}_0) := (H\omega, -\omega_0)$, and $(\gamma, \gamma_0) \in P(x^*, \eta^*)$ be maximal with respect to the objective $(\tilde{\omega}, \tilde{\omega}_0)$ such that $\tilde{\omega}^\top \gamma + \tilde{\omega}_0 \gamma_0 < 0$. Then the inequality $\gamma^\top Hx - \gamma_0 \eta \leq \gamma^\top b$ supports $\text{epi}(z)$.*

Proof Let $(\pi, \pi_0) := (H^\top \gamma, -\gamma_0)$. Then, by Remark 1.1, the statement is true if γ minimizes $\gamma^\top b$ among all possible certificates for the vector (π, π_0) in Theorem 1.2. It is easy to verify that γ is indeed a valid certificate for (π, π_0) in Theorem 1.2. For a contradiction, we hence assume that it does not minimize $\gamma^\top b$. Let thus $\gamma' \geq 0$ be an alternative certificate for (π, π_0) with $\gamma'^\top b < \gamma^\top b$. Then from (1.7) to (1.10) we obtain that $\gamma'^\top A - \pi_0 d^\top = 0$ and $\gamma'^\top H = \pi^\top$. Furthermore, since $(\gamma, \gamma_0) \in P^\leq(x^*, \eta^*)$,

$$\gamma'^\top (b - Hx^*) + \gamma_0 \eta^* = \gamma'^\top b - \pi^\top x^* + \gamma_0 \eta^* < \gamma^\top b - \pi^\top x^* + \gamma_0 \eta^* \leq -1.$$

We can thus scale (γ', γ_0) by an appropriate factor $\lambda \in (0, 1)$ to obtain that $\lambda \cdot (\gamma', \gamma_0) \in P^\leq(x^*, \eta^*)$ and

$$\tilde{\omega}^\top (\lambda \gamma') + \tilde{\omega}_0 \cdot (\lambda \gamma_0) = \lambda \cdot (\tilde{\omega}^\top \gamma' + \tilde{\omega}_0 \gamma_0) = \lambda \cdot (\tilde{\omega}^\top \gamma + \tilde{\omega}_0 \gamma_0) > \tilde{\omega}^\top \gamma + \tilde{\omega}_0 \gamma_0.$$

On the other hand, by Remark 2.1, if (γ, γ_0) maximizes the objective $(\tilde{\omega}, \tilde{\omega}_0)$ over $P(x^*, \eta^*)$, then it is also maximal within $P^\leq(x^*, \eta^*)$, a contradiction. By Remark 1.1, this implies that the inequality $\gamma^\top Hx - \gamma_0 \eta \leq \gamma^\top b$ does indeed support $\text{epi}(z)$. \square

A critical requirement for Corollary 2.3 is that $\tilde{\omega}^\top \gamma + \tilde{\omega}_0 \gamma_0 < 0$. Cornuéjols and Lemaréchal (2006, Theorem 2.3) establish some criteria on the objective function for which optimization problems over the reverse polar set are bounded. We have simplified the notation for our purposes and rephrased the relevant parts of the theorem according to our terminology.

Theorem 2.2 (Cornuéjols and Lemaréchal (2006)) *Let $(x^*, \eta^*) \notin \text{epi}(z)$, $(\omega, \omega_0) \in \mathbb{R}^n \times \mathbb{R}$, and*

$$z^* := \max \left\{ \omega^\top \pi + \omega_0 \pi_0 \mid (\pi, \pi_0) \in (\text{epi}(z) - (x^*, \eta^*))^- \right\}.$$

Then

$$z^* \begin{cases} \leq 0 & \text{if } (\omega, \omega_0) \in \text{cl}(\text{pos}(\text{epi}(z) - (x^*, \eta^*))) \\ = \infty & \text{otherwise.} \end{cases}$$

Furthermore, if $(\omega, \omega_0) \in (\text{epi}(z) - (x^, \eta^*))^-$, then $z^* \leq -1$.*

Note in particular that the last part of the above statement implies $z^* < 0$ whenever $(\omega, \omega_0) \in \text{pos}(\text{epi}(z) - (x^*, \eta^*)) \setminus \{0\}$, which provides us with a large variety of objective functions for which $\omega^\top \gamma + \omega_0 \gamma_0 < 0$ holds in the optimal solution. By Corollaries 2.2 and 2.3, this means that the cut which results from maximizing these objectives over the reverse polar set is guaranteed to be supporting.

2.2 Alternative representations

We now derive an alternative representation of the optimization problem in (2.4), which will turn out to be much more useful in practice. For instance, the structure of the resulting problem will be very similar to the original subproblem, which makes it easy to use existing solution algorithms for the subproblem in a cut-generating program.

Cornuéjols and Lemaréchal (2006, Theorem 4.2) prove that linear optimization problems over the reverse polar set can be evaluated in terms of the support function of the original set (in our case $\text{epi}(z) - (x^*, \eta^*)$). This can also be applied to the alternative polyhedron, as mentioned (without proof) by Fischetti et al. (2010). The following lemma makes a statement similar to Cornuéjols and Lemaréchal (2006, Theorem 4.2), which is applicable to a wider range of settings. For the proof, we refer to Stursberg (2019, Theorem 3.20).

Lemma 2.1 (Cornuéjols and Lemaréchal (2006)) *Let $K \subseteq \mathbb{R}^n$ be a cone and $c_1, c_2 \in \mathbb{R}^n$. Consider the optimization problems*

$$\max \left\{ c_1^\top x \mid x \in K, c_2^\top x = -1 \right\} \tag{2.5}$$

and

$$\max \left\{ c_2^\top x \mid x \in K, c_1^\top x \geq 1 \right\}. \tag{2.6}$$

Then the following hold:

- (a) If x^* is an optimal solution for (2.5) with objective value $\xi > 0$, then $\frac{1}{\xi} \cdot x^*$ is an optimal solution for (2.6) with objective value $-\frac{1}{\xi}$.
- (b) Conversely, if x^* is an optimal solution for (2.6) with objective value $\xi < 0$, then $-\frac{1}{\xi} \cdot x^*$ is an optimal solution for (2.6) with objective value $-\frac{1}{\xi}$.

This lemma allows us to solve optimization problems of the form (2.4) by instead resorting to the optimization problem

$$\max_{\gamma, \gamma_0 \geq 0} \gamma^\top (Hx^* - b) - \gamma_0 \eta^* \tag{2.7}$$

$$\gamma^\top A + \gamma_0 d^\top = 0 \tag{2.8}$$

$$\tilde{\omega}^\top \gamma + \tilde{\omega}_0 \gamma_0 = -1. \tag{2.9}$$

Let $(\tilde{\omega}, \tilde{\omega}_0) \in \mathbb{R}^m \times \mathbb{R}$ and let (γ^*, γ_0^*) denote an optimal solution with value $\xi > 0$ for (2.7) to (2.9). Applying Lemma 2.1 with $c_1 := (Hx^* - b, -\eta^*)$, $c_2 := (\tilde{\omega}, \tilde{\omega}_0)$ and $K := \left\{ (\gamma, \gamma_0) \geq 0 \mid \gamma^\top A + \gamma_0 d^\top = 0 \right\}$, we obtain that $\frac{1}{\xi} \cdot (\gamma^*, \gamma_0^*)$ is an optimal solution with value $-\frac{1}{\xi}$ for (2.4).

The structural similarity of (2.7) to (2.9) and the original problem becomes more apparent when we consider the dual problem:

Corollary 2.4 *Let $(\tilde{\omega}, \tilde{\omega}_0) \in \mathbb{R}^m \times \mathbb{R}$ and (λ, x, y) be an optimal solution for the problem*

$$\min \lambda \tag{2.10}$$

$$Ay \leq b - Hx^* - \lambda \tilde{\omega} \tag{2.11}$$

$$d^\top y \leq \eta^* - \lambda \tilde{\omega}_0 \tag{2.12}$$

with $\lambda > 0$. Denote the corresponding dual solution by (γ, γ_0) . Then $\frac{1}{\lambda}(\gamma, \gamma_0)$ is an optimal solution for

$$\max \left\{ \tilde{\omega}^\top \gamma + \tilde{\omega}_0 \gamma_0 \mid (\gamma, \gamma_0) \in P^\leq(x^*, \eta^*) \right\}$$

with objective value $-\frac{1}{\lambda}$.

Note that, together with our observations in the context of the definition of the alternative polyhedron (1.5), this means in particular that

- (a) Whenever (2.10) to (2.12) has objective value 0, then the alternative polyhedron is empty and $(x^*, \eta^*) \in \text{epi}(z)$, and
- (b) Whenever (2.10) to (2.12) is feasible with (finite) objective value greater than 0, then (2.3) and (2.4) have objective values strictly less than 0, which means that the requirements for Remark 2.1 and Corollary 2.3 are satisfied.

Remark 2.2 If $(\tilde{\omega}, \tilde{\omega}_0) := (H\omega, -\omega_0)$, then the optimization problem (2.10) to (2.12) becomes

$$\min \lambda \tag{2.13}$$

$$Ay \leq b - H(x^* + \lambda \cdot \omega) \tag{2.14}$$

$$d^\top y \leq \eta^* + \omega_0 \lambda \tag{2.15}$$

The difference between the formulations from Corollary 2.4 and Remark 2.2 lies in how they relax the original problem: In (2.10) to (2.12), the relaxation works on the level of individual inequalities by relaxing their right-hand sides, whereas in (2.13) to (2.15) it works on the level of the master solution (x^*, η^*) , allowing us to choose a possibly more advantageous value for the vector x itself.

3 Cut selection

As we have seen in the previous section, Benders decomposition can be viewed as an instance of a classical cutting plane algorithm (Theorem 1.2). The Benders subproblem takes the role of the separation problem and the alternative polyhedron that is commonly used to select a Benders cut is a higher-dimensional representation of the reverse polar set, which characterizes all possible cut normals (Theorem 2.1).

Finally, Corollary 2.4 and Remark 2.2 show that selecting a cut normal by a linear objective over the reverse polar set or the alternative polyhedron can be interpreted as two different relaxations (2.10) to (2.12) and (2.13) to (2.15) of the original Benders feasibility subproblem (1.4). The former relaxation provides more flexibility with respect to the choice of parameters and coincides with the latter for a particular selection of the objective function.

Cut selection is one of four major areas of algorithmic improvements for Benders decomposition that recent work has focused on [see, e. g., the very extensive literature review by Rahmaniani et al. (2017)]. As a consequence, a number of selection criteria for Benders cuts have previously been explicitly proposed in the literature. Many of them also arise naturally from our discussion and analysis of the Benders decomposition algorithm above. We will first present these criteria in the way they typically appear in the literature and then link them to the reverse polar set and/or the alternative polyhedron.

3.1 Minimal infeasible subsystems

The work of Fischetti et al. (2010) is based on the premise that “one is interested in detecting a ‘minimal source of infeasibility’” whenever the feasibility subproblem (1.4) is empty. They hence suggest to generate Benders cuts based on Farkas certificates that correspond to *minimal infeasible subsystems (MIS)* of (1.4). Fischetti et al. (2010) empirically study the performance of MIS-cuts on a set of multi-commodity network design instances. Their results suggest that MIS-based cut selection outperforms the standard implementation of Benders decomposition by a factor of at least

2–3. Furthermore, this advantage increases substantially when focusing on harder instances (e.g. those which could not be solved by the standard implementation within 10 hours).

We define this criterion as follows:

Definition 3.1 Let z be defined as in (1.2) and let $(\pi, \pi_0) \in \mathbb{R}^n \times \mathbb{R}$. We say that (π, π_0) satisfies the MIS criterion if there exists $(\gamma, \gamma_0) \geq 0$ such that

- (a) $\pi = H^\top \gamma, \pi_0 = -\gamma_0$
- (b) The inequalities of (1.4) corresponding to the rows of H which are multiplied by the non-zero components of (γ, γ_0) in the equations in a) form a minimal infeasible subsystem of (1.4).

Note that we have defined the MIS criterion as a property of a normal vector, rather than a property of a cut. The reason for this is that the cut normal is the only relevant choice to make, given that an optimal right-hand side for each cut normal is provided by Corollary 2.3. Accordingly, we will call any cut with a normal vector that satisfies the MIS criterion a MIS-cut.

Gleeson and Ryan (1990) show that the set of (γ, γ_0) that appear in the above definition is exactly (up to homogeneity) the set of vertices of the alternative polyhedron:

Theorem 3.1 (Gleeson and Ryan (1990)) *Let $(x^*, \eta^*) \in \mathbb{R}^n \times \mathbb{R}$. For each vertex v of the (relaxed) alternative polyhedron (1.5), the set of constraints corresponding to the non-zero entries of v forms a minimal infeasible subsystem of (1.4). Conversely, for every minimal infeasible subsystem, there exists a vertex of the alternative polyhedron.*

This theorem immediately provides a characterization of cut normals which satisfy MIS in terms of the alternative polyhedron, which is also used in Fischetti et al. (2010). However, we can furthermore use Theorem 2.1, to transfer one direction of the characterization to the reverse polar set:

Corollary 3.1 *Let z be defined as in (1.2) and $(x^*, \eta^*) \in \mathbb{R}^n \times \mathbb{R}$. If there is a vertex (γ, γ_0) of $P(x^*, \eta^*)$ such that $(\pi, \pi_0) = (H^\top \gamma, -\gamma_0)$, then the vector (π, π_0) satisfies the MIS criterion. Furthermore, if (π, π_0) is a vertex of $(\text{epi}(z) - (x^*, \eta^*))^-$, then it satisfies the MIS criterion.*

Note that the reverse direction of the last sentence is generally not true, i.e. there might be minimal infeasible subsystems that do not correspond to vertices of the reverse polar set. As an example to illustrate this, as well as the above definition overall, consider Example 1.1 and specifically Fig. 3: Each of the vectors $\tilde{P}_1, \tilde{P}_2, \tilde{P}_3$ satisfies the MIS criterion, since they can be related by the equations $\pi = H^\top \gamma, \pi_0 = -\gamma_0$ to vertices of the alternative polyhedron, which can be identified with minimal infeasible subsystems of (1.4).

3.2 Facet-defining cuts

In cutting plane algorithms for polyhedra, facet-defining cuts are commonly considered to be very useful since they form the smallest family of inequalities which completely

describe (the convex hull of) the feasible solutions. A cutting-plane algorithm that can separate (distinct) facet inequalities in each iteration is not necessarily computationally efficient, but at least it is automatically guaranteed to terminate after a finite number of iterations. Also in practical applications, facet cuts have turned out to be extremely useful, e.g. in the context of branch-and-cut algorithms for integer programs such as the Traveling Salesman Problem. This is why the description of facet-defining inequalities has been a large and very active area of research for decades (see Balas 1975; Nemhauser and Wolsey 1988; Cook et al. 1998; Korte and Vygen 2008 and, as mentioned before, Conforti and Wolsey 2018).

Remember that a halfspace $H_{(\pi,\alpha)}^{\leq}$ is facet-defining for a set C if $C \subseteq H_{(\pi,\alpha)}^{\leq}$ and $H_{(\pi,\alpha)} \cap C$ contains $\dim(C)$ many affinely independent points. Analogously to the MIS criterion above, we define the FACET criterion for a normal vector in the context of Benders decomposition as follows:

Definition 3.2 Let z be defined as in (1.2) and $(\pi, \pi_0) \in \mathbb{R}^n \times \mathbb{R} \setminus \{0\}$. We say that (π, π_0) satisfies the FACET criterion if there exists $\alpha \in \mathbb{R}$ such that $H_{((\pi,\pi_0),\alpha)}^{\leq}$ is either facet-defining for $\text{epi}(z)$ or the corresponding hyperplane $H_{((\pi,\pi_0),\alpha)}$ contains $\text{epi}(z)$.

Note that, in deviation from the common definition of a facet-defining cut, the above definition requires that the halfspace supports *at least* $\dim(C)$ affinely independent points. In other words, in the case where $\text{epi}(z)$ is not full-dimensional, we also allow that $\text{epi}(z)$ is entirely contained in the hyperplane which represents the boundary of $H_{((\pi,\pi_0),\alpha)}^{\leq}$. In this situation, the comparison of different cut normals is inherently difficult: Since there is no clear way to tell if a cut supporting a facet of $\text{epi}(z)$ or one fully containing the set is the stronger cut, the FACET criterion captures arguably the strongest statement about a cut in relation to $\text{epi}(z)$ that we can make in general: In no case would we want to select a cut that supports *neither* a facet *nor* fully contains the set $\text{epi}(z)$.

On the other hand, by this definition, the “trivial” cut normal $(0, 0)$ would be facet-defining (since the hyperplane $H_{((0,0),0)}$ contains all of $\mathbb{R}^n \times \mathbb{R}$ and thus also $\text{epi}(z)$). Since this is not very useful, we have to exclude this choice explicitly and thus choose $(\pi, \pi_0) \in \mathbb{R}^n \times \mathbb{R} \setminus \{0\}$ in Definition 3.2.

For an example to illustrate the above definition, we refer to Fig. 4 where it is discussed together with the property of Pareto-optimality, which will be defined later.

The following result was proven by Cornuéjols and Lemaréchal (2006, Theorem 6.2), containing a minor error in the case where the set P is subdimensional.⁴ We therefore re-state a corrected version of the important parts below, a corresponding proof can be found in Stursberg (2019, Theorem 3.30).

Theorem 3.2 (Cornuéjols and Lemaréchal (2006)) *Let $P \subseteq \mathbb{R}^n$ be a polyhedron, $x^* \notin P$ and*

$$r := \begin{cases} \dim(P) - 1, & x^* \in \text{aff}(P) \\ \dim(P), & x^* \notin \text{aff}(P). \end{cases} \tag{3.1}$$

⁴ The result can also be derived from Balas (1998, Theorem 4.5) in the context of disjunctive cuts.

Then, there exists an x^* -separating halfspace with normal vector $\pi \neq 0$ supporting an r -dimensional face of P if and only if there exists a vertex π^* of $\text{lin}(P - x^*) \cap (P - x^*)^-$ and some $\lambda > 0$ such that $\lambda\pi \in \pi^* + \text{lin}(P - x^*)^\perp$.

Most notably, for the case where P is full-dimensional (i.e., $\text{dim}(P) = n$) the above theorem implies that there exists an x^* -separating halfspace with normal vector π supporting a facet of P if and only if there exists a vertex π^* of $(P - x^*)^-$ and $\lambda \geq 0$ such that $\lambda\pi = \pi^*$.

In this case, every cut generated from a vertex of the reverse polar set defines a facet of $\text{epi}(z)$. If an explicit \mathcal{H} -representation of the reverse polar set is available, we can thus easily obtain a facet-defining cut, e.g., by linear programming.

Note that since $P^\leq(x^*, \eta^*)$ is line-free (i.e. its lineality space, the maximal linear subspace it contains, is $\{0\}$), Theorem 2.1 implies that for every vertex of the reverse polar set there exists a vertex of the relaxed alternative polyhedron (and hence of the original alternative polyhedron) that leads to the same cut normal. In other words, if the normal of an x^* -separating halfspace satisfies the FACET criterion, then it also satisfies the MIS criterion.

On the other hand, Theorem 2.1 is not sufficient to guarantee that selecting a vertex of the alternative polyhedron yields a facet-defining cut: As Example 1.1 shows, a vertex of $P^\leq(x^*, \eta^*)$, is not necessarily mapped to a vertex of the reverse polar set under the transformation from Theorem 2.1. This exposes a useful hierarchy of subsets of the alternative polyhedron according to the properties of the cut normals which they yield: Selecting a vertex of the alternative polyhedron already guarantees that the resulting cut normal satisfies the MIS criterion and the points that lead to cut normals satisfying the FACET criterion constitute a subset of these vertices. The approach of selecting MIS-cuts may thus be viewed as a heuristic method to find FACET-cuts.

Although cuts satisfying the MIS criterion in general do not satisfy the FACET criterion, we can obtain some information on when this is the case in the situation of Corollary 2.2, i.e. if the objective function $(\tilde{\omega}, \tilde{\omega}_0)$ used to select the cut via problem (2.4) satisfies $(\tilde{\omega}, \tilde{\omega}_0) = (H\omega, -\omega_0)$ for some valid objective (ω, ω_0) for problem (2.3).

In this case it turns out that we actually obtain a FACET-cut for all objectives (ω, ω_0) except those from a lower-dimensional subspace. More precisely, we can prove the following characterization of the relationship between vertices of the alternative polyhedron and cut normals satisfying the FACET criterion. Like Conforti and Wolsey (2018, Proposition 6), the following theorem provides a method to generate facet-defining cuts using a single linear program. In contrast to Conforti and Wolsey (2018), however, our theorem uses a linear program over the (relaxed) alternative polyhedron and thus creates a link to well-established cut selection methods in the context of Benders decomposition, such as that proposed by Fischetti et al. (2010):

Theorem 3.3 *Let z be defined as in (1.2), $(x^*, \eta^*) \in (\mathbb{R}^n \times \mathbb{R}) \setminus \text{epi}(z)$, and $(\omega, \omega_0) \in \text{cl}(\text{pos}(\text{epi}(z) - (x^*, \eta^*)))$. Then, there exists an optimal vertex $(\gamma^*, \gamma_0^*) \in P^\leq(x^*, \eta^*)$ with respect to the objective function $(H\omega, -\omega_0)$ such that the resulting cut normal $(H^\top \gamma^*, -\gamma_0^*)$ is (x^*, η^*) -separating and satisfies the FACET criterion.*

Proof Let $L := \text{lin}(\text{epi}(z) - (x^*, \eta^*))$ and observe that L is orthogonal to the lineality space of $(\text{epi}(z) - (x^*, \eta^*))^-$. From Theorem 2.2, the reverse polar set $(\text{epi}(z) - (x^*, \eta^*))^-$ is bounded in the direction of (ω, ω_0) . We may therefore choose an optimal solution (π, π_0) from the intersection $(\text{epi}(z) - (x^*, \eta^*))^- \cap L$. While the reverse polar need not be line-free, note that $(\text{epi}(z) - (x^*, \eta^*))^- \cap L$ is indeed line-free and we can therefore choose (π, π_0) to be extremal in $(\text{epi}(z) - (x^*, \eta^*))^- \cap L$. By Corollary 2.2, there exists γ' with $H^\top \gamma' = \pi$ such that $(\gamma', -\pi_0)$ is an optimal solution to the problem

$$\max \left\{ (H\omega)^\top \gamma - \omega_0 \gamma_0 \mid (\gamma, \gamma_0) \in P^\leq(x^*, \eta^*) \right\}. \tag{3.2}$$

Denote by P^* the face of optimal solutions of (3.2) and observe that

$$\begin{aligned} (\gamma', -\pi_0) \in P^* \cap \left\{ (\gamma, \gamma_0) \mid (H^\top \gamma, -\gamma_0) - (\pi, \pi_0) = 0 \right\} \\ \subseteq P^* \cap \left\{ (\gamma, \gamma_0) \mid (H^\top \gamma, -\gamma_0) - (\pi, \pi_0) \in L^\perp \right\}. \end{aligned}$$

Let (γ^*, γ_0^*) be a vertex of $P^* \cap \left\{ (\gamma, \gamma_0) \mid (H^\top \gamma, -\gamma_0) - (\pi, \pi_0) \in L^\perp \right\}$ (which exists, since $P^\leq(x^*, \eta^*)$ is line-free). Then (γ^*, γ_0^*) is obviously optimal for (3.2). Since furthermore $(H^\top \gamma^*, -\gamma_0^*) = (\pi, \pi_0) + v$ for some $v \in L^\perp$, this means by Theorem 3.2 that it satisfies the FACET criterion.

It remains to show that (γ^*, γ_0^*) is a vertex of P^* , thus showing that it is also a vertex of $P^\leq(x^*, \eta^*)$. To see this, let $(\gamma^1, \gamma_0^1), (\gamma^2, \gamma_0^2) \in P^*$ such that $(\gamma^*, \gamma_0^*) \in \text{relint}([\gamma^1, \gamma_0^1), (\gamma^2, \gamma_0^2)])$ ⁵ However, it follows that

$$(\pi, \pi_0) + v = (H^\top \gamma^*, -\gamma_0^*) \in \text{relint}([(H^\top \gamma^1, -\gamma_0^1), (H^\top \gamma^2, -\gamma_0^2)])$$

and by Theorem 2.1, $[(H^\top \gamma^1, -\gamma_0^1), (H^\top \gamma^2, -\gamma_0^2)] \subseteq (\text{epi}(z) - (x^*, \eta^*))^-$. As (π, π_0) is extremal in the set $(\text{epi}(z) - (x^*, \eta^*))^- \cap L$, this implies that both $(H^\top \gamma^1, -\gamma_0^1), (H^\top \gamma^2, -\gamma_0^2) \in \left\{ (\pi, \pi_0) + v \mid v \in L^\perp \right\}$ which in turn implies that $(\gamma^1, \gamma_0^1), (\gamma^2, \gamma_0^2) \in P^* \cap \left\{ (\gamma, \gamma_0) \mid (H^\top \gamma, -\gamma_0) - (\pi, \pi_0) \in L^\perp \right\}$. As (γ^*, γ_0^*) is extremal in $P^* \cap \left\{ (\gamma, \gamma_0) \mid (H^\top \gamma, -\gamma_0) - (\pi, \pi_0) \in L^\perp \right\}$, we obtain that $(\gamma^1, \gamma_0^1) = (\gamma^2, \gamma_0^2) = (\gamma^*, \gamma_0^*)$, which proves extremality of (γ^*, γ_0^*) in P^* . \square

In particular, Theorem 3.3 implies the following: If $(\gamma', \gamma'_0) \in P^\leq(x^*, \eta^*)$ is an optimal vertex with respect to the objective function $(H\omega, -\omega_0)$ such that the resulting cut normal $(H^\top \gamma', -\gamma'_0)$ does not satisfy the FACET criterion, then the optimal solution for maximizing $(H\omega, -\omega_0)$ over $P^\leq(x^*, \eta^*)$ is not unique. Furthermore, by

⁵ $\text{relint}(S)$ ($\text{relbd}(S)$) denote the interior (boundary) relative to the affine hull of S .

Corollary 2.2, this implies that the same is true for maximizing the objective (ω, ω_0) over $\text{epi}(z) - (x^*, \eta^*)^-$.

We can summarize our results as follows: While any FACET-cut is also an MIS-cut, the reverse is not always true. However, if we optimize the objective $(H\omega, -\omega_0)$ over the alternative polyhedron, then there exists only a subdimensional set of choices for the vector (ω, ω_0) for which the resulting cut might not satisfy the FACET criterion (those, for which the optimum over the reverse polar set is non-unique).

This suggests that these cases should be “rare” in practice, especially if we choose (or perturb) (ω, ω_0) randomly from some full-dimensional set. This argument, why a cut obtained for a generic vector (ω, ω_0) can be expected to be facet-defining, is identical to the concept of “almost surely” finding facet-defining cuts proposed by Conforti and Wolsey (2018).

Looking back at Remark 2.2, this similarity should not come as a surprise: With $(\omega, \omega_0) = (\bar{x} - x^*, \bar{\eta} - \eta^*)$ for a point $(\bar{x}, \bar{\eta}) \in \text{relint}(\text{epi}(z))$, the resulting cut-generating LP is almost identical. In fact, the point $(\bar{x}, \bar{\eta})$ in this case takes the role of the point that the origin is relocated into in the approach from Conforti and Wolsey (2018). Observe, however, that while Conforti and Wolsey (2018) require that point to lie in the relative interior of $\text{epi}(z)$, we can actually expect a cut satisfying the FACET criterion from any (ω, ω_0) for which the optimal objective over the reverse polar is strictly negative. By Theorem 2.2, one sufficient (but not necessary) criterion for this is to choose $(\omega, \omega_0) = (\bar{x} - x^*, \bar{\eta} - \eta^*)$ for an arbitrary point $(\bar{x}, \bar{\eta}) \in \text{relint}(\text{epi}(z))$.

3.3 Pareto-optimality

The first systematic work on the general selection of Benders cuts to our knowledge was undertaken by Magnanti and Wong (1981). The paper, which has proven very influential and still being referred to regularly, focuses on the property of Pareto-optimality. It can intuitively be described as follows: A cut is Pareto-optimal if there is no other cut valid for $\text{epi}(z)$ which is *clearly superior*, which *dominates* the first cut.

In this setting, any cut that does not support $\text{epi}(z)$ is obviously dominated. Between supporting cuts, there is no general criterion for domination. We can, however, compare cuts where the cut normal (π, π_0) satisfies $\pi_0 \neq 0$ (this is also the case covered by Magnanti and Wong (1981)):

Definition 3.3 For a problem of the form (1.1), we say that an inequality $\pi^\top x + \pi_0 \eta \leq \alpha$ with $\pi_0 < 0$ is *dominated* by another inequality $\pi'^\top x + \pi'_0 \eta \leq \alpha'$ if $\pi'_0 < 0$ and

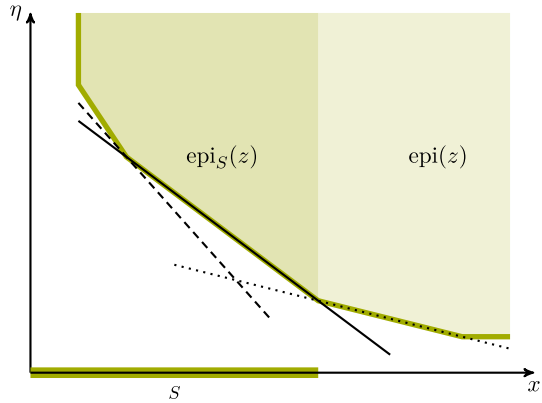
$$\frac{\pi'^\top x - \alpha'}{-\pi'_0} \geq \frac{\pi^\top x - \alpha}{-\pi_0} \quad \text{for all } x \in S, \tag{3.3}$$

with strict inequality for at least one $x \in S$.

If $\pi_0 < 0$ and $\pi^\top x + \pi_0 \eta \leq \alpha$ is not dominated by any valid inequality for $\text{epi}(z)$, then we call it *Pareto-optimal*.

Remember that the set S contains all points $x \in \mathbb{R}^n$ that are feasible for an optimization problem of the form (1.1) if we ignore the linear constraints $Hx + Ay \leq b$.

Fig. 4 The dotted cut supports a facet of $\text{epi}(z)$ and it supports $\text{epi}_S(z)$, but it is still not Pareto-optimal. The solid cut supports a facet of $\text{epi}_S(z)$ and is hence Pareto-optimal. The dashed cut is Pareto-optimal even though it does not support a facet of $\text{epi}(z)$ (or $\text{epi}_S(z)$)



By the above definition, a cut dominates another cut if the minimum value of η that it enforces is at least as good for all $x \in S$ and strictly better for at least one $x \in S$ (see Fig. 4).

Analogously to the previous criteria, we define the PARETO criterion for a cut normal:

Definition 3.4 For a problem of the form (1.1) with z as defined as in (1.2), let $(\pi, \pi_0) \in \mathbb{R}^n \times \mathbb{R}$. We say that (π, π_0) satisfies the PARETO criterion if there exists a scalar $\alpha \in \mathbb{R}$ such that the inequality $\pi^\top x + \pi_0 \eta \leq \alpha$ is Pareto-optimal.

This criterion is very reasonable: If a cut is not Pareto-optimal, then there exists a different cut which is also valid for $\text{epi}(z)$, but leads to a strictly tighter approximation. We would hence prefer to generate a stronger, Pareto-optimal cut right away.

The following theorem provides us with a characterization of Pareto-optimal cuts. It is based on the idea of Magnanti and Wong (1981, Theorem 1), which is formulated under the assumption that the subproblem is always feasible (which implies that $\pi_0 < 0$ for any cut normal (π, π_0)). While the original theorem is only concerned with sufficiency, we extend the result in a natural way to obtain a criterion that gives a complete characterization of Pareto-optimal cuts. We use the following separation lemma:

Lemma 3.1 (Rockafellar (1970)) Let $C \subseteq \mathbb{R}^n$ be a non-empty convex set and $K \subseteq \mathbb{R}^n$ a non-empty polyhedron such that $\text{relint}(C) \cap K = \emptyset$. Then, there exists a hyperplane separating C and K which does not contain C .

Using this lemma, we obtain the following theorem:

Theorem 3.4 For a problem of the form (1.1), let $(\pi, \pi_0) \in \mathbb{R}^n \times \mathbb{R}$ with $\pi_0 < 0$. The inequality $\pi^\top x + \pi_0 \eta \leq \alpha$ is Pareto-optimal if and only if $H_{((\pi, \pi_0), \alpha)}^\leq$ is a halfspace supporting $\text{epi}(z)$ in a point $(x^*, \eta^*) \in \text{epi}(z) \cap \text{relint}(\text{conv}(S)) \times \mathbb{R}$.

Proof For the *if* part, suppose for a contradiction that $(\pi^\top, \pi_0)(x, \eta)^\top \leq \alpha$ is not Pareto-optimal, i.e. there exist some π', π'_0, α' such that the inequality

$(\pi'^\top, \pi'_0)(x, \eta)^\top \leq \alpha'$ dominates the former inequality. This means that for all $x \in S$ (and hence all $x \in \text{conv}(S)$), it holds that

$$\frac{\pi'^\top x - \alpha'}{-\pi'_0} \geq \frac{\pi^\top x - \alpha}{-\pi_0} \tag{3.4}$$

and furthermore

$$\frac{\pi'^\top \bar{x} - \alpha'}{-\pi'_0} > \frac{\pi^\top \bar{x} - \alpha}{-\pi_0} \quad \text{for some } \bar{x} \in S.$$

Finally, since $H_{((\pi, \pi_0), \alpha)}^{\leq}$ supports $\text{epi}(z)$ in (x^*, η^*) ,

$$\frac{\pi'^\top x^* - \alpha'}{-\pi'_0} \leq \eta^* = \frac{\pi^\top x - \alpha}{-\pi_0} \leq \frac{\pi'^\top x^* - \alpha'}{-\pi'_0}$$

and hence equality must hold everywhere in the above inequality chain. Now, as $x^* \in \text{relint}(\text{conv}(S))$, we can choose $\lambda > 1$ such that $\tilde{x} := \bar{x} + \lambda(x^* - \bar{x}) \in \text{conv}(S)$. But then

$$\begin{aligned} \frac{\pi'^\top \tilde{x} - \alpha'}{-\pi'_0} &= \underbrace{(1 - \lambda)}_{<0} \underbrace{\frac{\pi'^\top \bar{x} - \alpha'}{-\pi'_0}}_{> \frac{\pi^\top \bar{x} - \alpha}{-\pi_0}} + \lambda \underbrace{\frac{\pi'^\top x^* - \alpha'}{-\pi'_0}}_{= \frac{\pi^\top \bar{x} - \alpha}{-\pi_0}} \\ &< \frac{\pi^\top \tilde{x} - \alpha}{-\pi_0}, \end{aligned}$$

contradicting (3.4).

For the *only-if* part, we first note that if $H_{((\pi, \pi_0), \alpha)}^{\leq}$ does not support $\text{epi}(z)$, then it is obviously dominated by $H_{((\pi, \pi_0), \alpha')}^{\leq}$ with $\alpha' := \alpha + \varepsilon$ for some $\varepsilon > 0$. Therefore, let $H_{((\pi, \pi_0), \alpha)}^{\leq}$ be such that it supports $\text{epi}(z)$, but not in points from the set $\text{epi}(z) \cap \text{relint}(\text{conv}(S)) \times \mathbb{R}$. Denote by $S^* := \left\{ x \in \mathbb{R}^n \mid \exists \eta : (x, \eta) \in \text{epi}(z) \cap H_{((\pi, \pi_0), \alpha)} \right\}$ the set of points where $H_{((\pi, \pi_0), \alpha)}^{\leq}$ supports $\text{epi}(z)$.

Since $\text{relint}(\text{conv}(S)) \cap S^* = \emptyset$, we can use Lemma 3.1 to obtain a hyperplane separating $\text{conv}(S)$ and S^* which does not contain S . Hence, there exist π^*, α^* such that $\pi^{*\top} x \geq \alpha^*$ for all $x \in S^*$ and $\pi^{*\top} x \leq \alpha^*$ for all $x \in \text{conv}(S)$, where the second inequality is strict for some $x \in \text{conv}(S)$ and thus also for some $x^* \in S$.

Let $\varepsilon > 0$, $\pi' := \pi - \varepsilon\pi^*$ and $\alpha' := \alpha - \varepsilon\alpha^*$. If ε is sufficiently small, then the inequality $(\pi'^\top, \pi_0)(x, \eta)^\top \leq \alpha'$ is valid for $\text{epi}(z)$: All $(x, \eta) \in \text{epi}(z)$ with $x \notin S^*$ satisfy the original inequality strictly and for all $x \in S^*$,

$$(\pi'^\top, \pi_0)(x, \eta)^\top - \alpha' = (\pi^\top, \pi_0)(x, \eta)^\top - \alpha - \varepsilon \underbrace{(\pi^{*\top} x - \alpha^*)}_{\geq 0}$$

$$\leq (\pi^\top, \pi_0)(x, \eta)^\top - \alpha \leq 0,$$

since the original inequality is valid for $\text{epi}(z)$.

Finally, we claim that the inequality $(\pi^\top, \pi_0)(x, \eta)^\top \leq \alpha$ is dominated by the inequality $(\pi'^\top, \pi_0)(x, \eta)^\top \leq \alpha'$: For all $x \in S$, it holds that

$$\frac{\pi'^\top x - \alpha'}{-\pi_0} = \frac{\pi^\top x - \alpha}{-\pi_0} + \varepsilon \cdot \frac{\pi^{*\top} x - \alpha^*}{-\pi_0} \geq \frac{\pi^\top x - \alpha}{-\pi_0}.$$

Since the last inequality is strict for $x^* \in S$, this proves the statement. □

For the case where S is convex, the previous theorem immediately implies the following statement:

Corollary 3.2 *Let S be convex. Then, $\pi^\top x + \pi_0 \eta \leq \alpha$ is Pareto-optimal if and only if $H_{((\pi, \pi_0), \alpha)}^{\leq}$ supports a face F of $\text{epi}_S(z)$ such that $F \not\subset \text{reld}(S) \times \mathbb{R}$.*

Magnanti and Wong (1981) also propose an algorithm that computes a Pareto-optimal cut by solving the cut-generating problem twice. While their algorithm is defined for the original Benders optimality cuts, it can be adapted to work with other cut selection criteria, as well. Sherali and Lunday (2013) present a method based on multiobjective optimization to obtain a cut that satisfies a weaker version of Pareto-optimality by solving only a single instance of the cut-generating LP. Papadakos (2008) notes that, given a point in the relative interior of $\text{conv}(S)$, a Pareto-optimal cut can be generated using a single run of the cut-generating problem. Also, under certain conditions on the problem, other points not in the relative interior allow this, as well. However, the approach suggested by the authors adds Pareto-optimal cuts independently from master- or subproblem solutions, together with subproblem-generated cuts, which are generally not Pareto-optimal. This means that the Pareto-optimal cuts which are added may not even cut off the current tentative solution. The upcoming Theorem 3.5 will lead to an approach that reconciles both objectives, generating cuts that are both Pareto-optimal and cut off the current tentative solution.

We use a result by Cornuéjols and Lemaréchal (2006) on the set of points exposed by a cut normal (π, π_0) to derive a method that always obtains a Pareto-optimal cut. The following lemma has been slightly generalized and rewritten to match our setting and notation, but it follows the general idea of Cornuéjols and Lemaréchal (2006, Theorem 3.4).

Lemma 3.2 (Cornuéjols and Lemaréchal (2006)) *Let $(x^*, \eta^*) \in \mathbb{R}^n \times \mathbb{R} \setminus \text{epi}(z)$ and $(\omega, \omega_0) \in \text{pos}(\text{epi}(z) - (x^*, \eta^*))$ and let*

$$(\bar{x}, \bar{\eta}) := \frac{(\omega, \omega_0)}{-h_Q(\omega, \omega_0)} + (x^*, \eta^*).$$

where $h_Q(\omega, \omega_0) := \sup \left\{ \omega^\top x + \omega_0 \eta \mid (x, \eta) \in Q \right\}$ is the support function of the set Q .

Table 1 Properties of a cut resulting from a vertex in the alternative polyhedron which maximizes $(\tilde{\omega}, \tilde{\omega}_0)$ (under the assumption that a finite optimum exists)

$(\tilde{\omega}, \tilde{\omega}_0)$	MIS	FACET	PARETO
$\in \mathbb{R}^n \times \mathbb{R}$	✓	✗	✗
$\in (H, -1) \cdot \mathbb{R}^n \times \mathbb{R}$	✓	(✓)	✗
$\in (H, -1) \cdot \text{relint}(\text{conv}(\text{epi}_S(z) - (x^*, \eta^*)))$	✓	(✓)	✓

The checkmark in parentheses (✓) indicates that the property is satisfied for all $(\tilde{\omega}, \tilde{\omega}_0)$ in the specified set except those from a specific sub-dimensional subset

If (π, π_0) is optimal in $Q := (\text{epi}(z) - (x^*, \eta^*))^-$ with respect to the objective (ω, ω_0) then there exists $\alpha \in \mathbb{R}$ such that $H_{((\pi, \pi_0), \alpha)}^{\leq}$ supports $\text{epi}(z)$ in $(\bar{x}, \bar{\eta})$.

Proof The case of $(\omega, \omega_0) \in (\text{epi}(z) - (x^*, \eta^*))$ was proven by Cornuéjols and Lemaréchal (2006, Theorem 3.4). If $(\omega, \omega_0) \in \text{pos}(\text{epi}(z) - (x^*, \eta^*))$, then there is $\mu > 0$ such that $\mu \cdot (\omega, \omega_0) \in (\text{epi}(z) - (x^*, \eta^*))$. Note that if (π, π_0) is optimal with respect to (ω, ω_0) , then also with respect to $\mu \cdot (\omega, \omega_0)$. Thus it follows from Cornuéjols and Lemaréchal (2006, Theorem 3.4) that there exists $\alpha \in \mathbb{R}$ such that $H_{(\pi, \pi_0), \alpha}^{\leq}$ supports $\text{epi}_S(z)$ in

$$(\bar{x}, \bar{\eta}) := \frac{\mu \cdot (\omega, \omega_0)}{-h_Q(\mu\omega, \mu\omega_0)} + (x^*, \eta^*) = \frac{(\omega, \omega_0)}{-h_Q(\omega, \omega_0)} + (x^*, \eta^*).$$

□

We can now prove the theorem already mentioned above.

Theorem 3.5 Let $(x^*, \eta^*) \in S \times \mathbb{R}$, $(\omega, \omega_0) \in \text{relint}(\text{conv}(\text{epi}_S(z) - (x^*, \eta^*)))$, (π, π_0) be optimal in $(\text{epi}(z) - (x^*, \eta^*))^-$ with respect to the objective (ω, ω_0) , and $\pi_0 < 0$. Then (π, π_0) satisfies the PARETO criterion.

Proof Let $Q := (\text{epi}(z) - (x^*, \eta^*))^-$ again and $\lambda := -(h_Q(\omega, \omega_0))^{-1}$. Since, in particular, $(\omega, \omega_0) \in \text{epi}_S(z) - (x^*, \eta^*)$ it follows from the definition of the reverse polar set that $h_Q(\omega, \omega_0) \leq -1$ and thus $\lambda \in [0, 1]$.

For $(\bar{x}, \bar{\eta})$ from Lemma 3.2, we obtain $(\bar{x}, \bar{\eta}) = \lambda((\omega, \omega_0) + (x^*, \eta^*)) + (1 - \lambda)(x^*, \eta^*)$ is a convex combination of $(\omega, \omega_0) + (x^*, \eta^*) \in \text{relint}(\text{conv}(\text{epi}_S(z))) \subseteq \text{relint}(\text{conv}(S)) \times \mathbb{R}$ and $(x^*, \eta^*) \in S \times \mathbb{R}$. Therefore, $\bar{x} \in \text{relint}(\text{conv}(S))$ and thus by Theorem 3.4 the cut defined by (π, π_0) is Pareto-optimal. □

The results from this section are summarized in Table 1.

4 Computational results

To validate the theoretical results presented in this paper, we have compared our refined cut selection approach to that presented by Fischetti et al. (2010) on a set of instances of the Capacity Expansion Problem for electrical power systems.

To map the range of potential instances we use a total of 14 test instances, spanning from a small, closely connected model of the Bavarian power system to a large, realistic model of the (rather sparsely connected) European power system consisting of 102 demand regions with 587 (aggregated) generation units and 195 existing and potential transmission lines (Schaber et al. 2012), that were investigated in the context of two joint research projects.

For both models, we optimize capacity expansion and hourly dispatch based on demand data and data for the availability of renewable energy sources in hourly resolution for a period of one year. Due to their inherent structure, where subproblems for individual timesteps are loosely coupled by capacity expansion decisions and storage constraints, this type of problem is generally well-suited for Benders decomposition. To give an indication of the size of the resulting optimization problems, instances 1-12 each consist of $\approx 800,000$ variables, $\approx 1,200,000$ constraints and $\approx 3,700,000$ non-zero entries in the constraint matrix. Instances 13 and 14 consist of $\approx 1,500,000$ variables, $\approx 2,600,000$ constraints and $\approx 7,600,000$ non-zero entries in the constraint matrix.

To demonstrate the benefits of our refined cut selection approach, we continuously update the weight vector $(\tilde{\omega}, \tilde{\omega}_0)$ so that in each iteration it satisfies the conditions of Theorem 3.5 and thus the resulting cut meets the most advantageous of the criteria developed in this paper (corresponding to the last row in Table 1). We call this approach *adaptive cuts*.

As a benchmark, we use a version of the approach proposed by Fischetti et al. (2010) that strengthens the resulting cuts without additional computational effort using the information represented by the matrix H (thereby in particular making sure that the obtained cut is always supporting). We denote this approach by the term *static cuts*.

More specifically, we compare the following approaches:

“adaptive cuts” $(\tilde{\omega}, \tilde{\omega}_0) = (H, -1)^\top (\bar{x} - x^*, \bar{\eta} - \eta^*)$ where $(\bar{x}, \bar{\eta})$ is a (suboptimal) feasible solution computed based on a current upper bound from the Benders Decomposition algorithm.

“static cuts” $(\tilde{\omega}, \tilde{\omega}_0) = (H, -1)^\top (\mathbb{1}, 1)$.

As mentioned above, the computational effort for the cut-generating LP in both approaches is almost identical: The only difference is that in the “adaptive cuts” approach, we use a different objective function, which can be obtained from the result of the previous iteration via a simple matrix-vector multiplication.

We have implemented both approaches in C++ using Gurobi 7.5. For our computations, we used ten CPU cores running at 2.4 GHz with 45 GB of main memory.

To solve master problem and subproblems, we use the dual simplex algorithm (i. e., the version of the simplex algorithm that maintains dual feasibility while pivoting between bases). In each iteration, we warm-start all problems using the optimal basis from the previous iteration and solve all subproblems in parallel on the available cores. Beyond this, we run the solution algorithm with default settings, i. e., we did not undertake any computational optimizations with respect to either the algorithm itself or the solution method of master and subproblems. In particular, also the specific update mechanism use for the *adaptive cuts* approach should be seen as an illustrative example rather than a performance-optimized prescription.

Table 2 Time (in hours) to reach different relative duality gaps with *static cuts* and *adaptive cuts*, respectively, for different instances in our test set

Gap	Instance 1		Instance 2		Instance 3		Instance 4	
	Static	Adaptive	Static	Adaptive	Static	Adaptive	Static	Adaptive
1.00×10^{-1}	3.10	1.37	2.90	1.44	2.99	1.15	3.20	1.07
1.00×10^{-2}	6.61	2.20	5.36	1.94	6.53	1.86	6.02	1.70
1.00×10^{-3}	10.11	3.60	8.57	3.16	9.90	3.05	8.81	2.69
1.00×10^{-4}	13.46	5.04	12.05	4.24	13.83	4.47	11.72	3.56
1.00×10^{-5}	15.69	6.06	14.45	5.28	17.37	5.69	14.60	4.82
Gap	Instance 5		Instance 6		Instance 7		Instance 8	
	Static	Adaptive	Static	Adaptive	Static	Adaptive	Static	Adaptive
1.00×10^{-1}	2.99	1.22	3.11	1.08	2.87	1.22	3.34	1.32
1.00×10^{-2}	5.29	2.00	5.52	1.68	5.17	1.99	5.58	2.57
1.00×10^{-3}	8.91	3.52	9.13	2.44	8.69	3.05	8.52	4.10
1.00×10^{-4}	13.13	5.10	13.58	3.66	13.88	4.67	12.38	5.74
1.00×10^{-5}	17.52	6.99	18.06	5.39	18.41	6.21	15.49	6.86
Gap	Instance 9		Instance 10		Instance 11		Instance 12	
	Static	Adaptive	Static	Adaptive	Static	Adaptive	Static	Adaptive
1.00×10^{-1}	2.48	1.24	2.68	1.08	3.08	1.32	2.98	1.31
1.00×10^{-2}	5.79	2.01	6.13	1.95	6.32	2.46	6.16	2.09
1.00×10^{-3}	9.41	3.12	9.64	3.17	9.55	3.51	8.88	3.51
1.00×10^{-4}	14.01	4.29	13.59	4.31	12.89	4.53	11.69	4.68
1.00×10^{-5}	17.76	5.86	18.28	6.23	15.10	5.34	13.98	5.44
Gap	Instance 13		Instance 14					
	Static	Adaptive	Static	Adaptive				
1.00×10^{-1}	2.60	1.46	2.38	1.25				
1.00×10^{-2}	2.85	1.50	2.53	1.25				
1.00×10^{-3}	3.25	1.65	3.09	1.44				
1.00×10^{-4}	3.81	1.76	3.75	1.66				
1.00×10^{-5}	4.06	1.95	4.10	1.85				

As a performance measure, we use the time to reach different thresholds for the relative duality gap, i. e., the gap between upper and lower bound relative to the optimal objective value. This takes into account that in practical applications, one is often satisfied with a solution that is guaranteed to be within a certain tolerance of the optimal solution (e. g., 0.1 %), rather than a strictly optimal solution. Our results show that for any desired gap, the *adaptive cuts* selection approach performs substantially better than the *static cuts* approach.

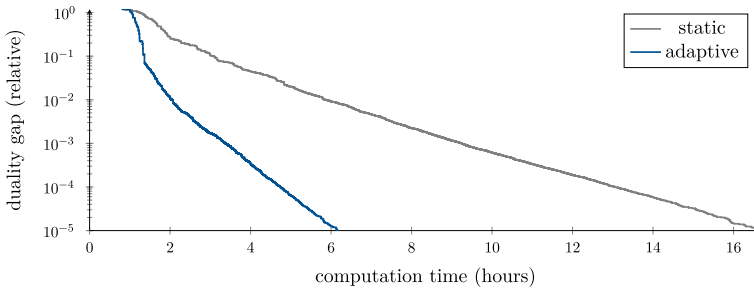


Fig. 5 Progression of the average relative duality gap using *static cuts* and *adaptive cuts* over all our test instances based on the European power system: The *adaptive cuts* reduce the gap by a factor of 2–3 faster compared to the *static cuts* approach

The results can be inspected in detail in Table 2: The *adaptive cuts* approach generally reaches any given optimality threshold by a factor of 2–3 faster than the *static cuts* approach. While the general result is very consistent across all instances, differences in the magnitude of the advantage exist: The benefit of the *adaptive cuts* approach tends to be larger in more difficult instances which overall take longer to solve.

To further visualize our results, Fig. 5 compares the two approaches with respect to the progression of the average duality gap over all instances based on the (larger) European power system model. The plot confirms our observation from Table 2: The *adaptive cuts* approach reduces the duality gap by a factor of 2–3 faster than the *static cuts* approach.

5 Outlook

We conclude with an outlook on interesting research questions raised by the results presented in this paper.

Our theoretical results clearly point to a computational advantage from improving the parametrization of cut-generating LPs in Benders decomposition and we have demonstrated this advantage in the context of Capacity Expansion Problems for electrical power systems. This holds despite the fact that we have not performed any “fine tuning” of parameters beyond what is immediately implied by our theoretical results. A broader computational study of such optimizations (for which we point out some ideas below), as well as of general performance across other types of problem instances would certainly be worthwhile.

In a generic implementation of Benders decomposition, feasible solutions are used primarily to decide when the algorithm has converged sufficiently close to the optimal solution. By Theorem 2.2, however, any such solution can furthermore be used to derive a subproblem objective which satisfies the prerequisites of both Corollary 2.3 and Theorem 3.3. Together with Theorem 3.5, they thus result in the generation of cuts which are always supporting, almost always support a facet and (if $\pi_0 < 0$) are also pareto-optimal. In our computational experiments, these cuts proved to be very useful in improving the performance of a Benders decomposition algorithm. Since

information from a feasible solution can thus be used within the cut generation, it makes sense to investigate more closely the possibilities how such a solution can be obtained during the algorithm. This is likely to be very problem-specific, but some general ideas could be:

- How is the information from feasible solutions computed in different iterations best aggregated? Does it make sense to use e. g. a stabilization approach or a convex combination with some other choices for (ω, ω_0) , e. g. from previous iterations? This corresponds to the method used by Papadakos (2008) in their empirical study.
- More broadly, what different methods can be used to generate feasible solutions and what effect do different feasible solutions have on cut generation and the computational performance of the algorithm?

Furthermore, if a feasible solution is not available as the basis for a subproblem objective, the cut-generating problem might be unbounded/infeasible. On the other hand, the approach from Fischetti et al. (2010) with $\tilde{\omega} = \mathbb{1}$ yields a cut-generating LP that is always feasible, but the resulting cut might be weaker. How can both approaches be combined in a best-possible way? For instance, is choosing $\tilde{\omega} = H\omega + \varepsilon \cdot \mathbb{1}$ as the relaxation term and letting ε go to zero a good choice?

Finally, our approach provides a clear geometric interpretation of the interaction between parametrization of the cut-generating LP and the resulting cut normals. How can this be used to leverage a-priori knowledge about the problem (or information obtained through a fast preprocessing algorithm) to improve the selection of a subproblem objective (ω, ω_0) from a set of cuts satisfying the same quality criteria (e.g. that are all facet-defining)?

Acknowledgements The authors thank their colleagues Magdalena Stüber and Michael Ritter for their invaluable support as well as the reviewers and the associate editor for their helpful comments and the corresponding discussion, which have contributed to further improve the paper. Paul Stursberg acknowledges funding from Deutsche Forschungsgemeinschaft (DFG) through TUM International Graduate School of Science and Engineering (IGSSE), GSC 81, and by the German Federal Ministry for Economic Affairs and Energy (FKZ 03ET4029) on the basis of a decision by the German Bundestag.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Balas E (1975) Facets of the knapsack polytope. *Math Program* 8(1):146–164
- Balas E (1998) Disjunctive programming: properties of the convex hull of feasible points. *Discrete Appl Math* 89(1):3–44
- Balas E, Ivanescu PL (1964) On the generalized transportation problem. *Manag Sci* 11(1):188–202

- Benders JF (1962) Partitioning procedures for solving mixed-variables programming problems. *Numer Math* 4(1):238–252
- Conforti M, Wolsey LA (2018) Facet separation with one linear program. *Math Program Ser A* 178:1–20
- Cook W, Cunningham W, Pullyblank W, Schrijver A (1998) *Combinatorial optimization. Series in discrete mathematics and optimization*. Wiley
- Cornuéjols G, Lemaréchal C (2006) A convex-analysis perspective on disjunctive cuts. *Math Program Ser A* 106(3):567–586
- Fischetti M, Salvagnin D, Zanette A (2010) A note on the selection of Benders cuts. *Math Program Ser B* 124(1–2):175–182
- Gleeson J, Ryan J (1990) Identifying minimally infeasible subsystems of inequalities. *ORSA J Comput* 2(1):61–63
- Korte B, Vygen J (2008) *Combinatorial optimization*, 4th edn. Springer
- Magnanti TL, Wong R (1981) Accelerating Benders decomposition: algorithmic enhancement and model selection criteria. *Oper Res* 29(3):464–484
- Nemhauser GL, Wolsey LA (1988) *Integer and combinatorial optimization*. Wiley
- Papadakos N (2008) Practical enhancements to the Magnanti-Wong method. *Oper Res Lett* 36(4):444–449
- Rahmaniani R, Crainic TG, Gendreau M, Rei W (2017) The Benders decomposition algorithm: a literature review. *Eur J Oper Res* 259(3):801–817
- Rockafellar RT (1970) *Convex analysis*. Princeton University Press
- Schaber K, Steinke F, Hamacher T (2012) Transmission grid extensions for the integration of variable renewable energies in Europe: Who benefits where? *Energy Policy* 43:123–135
- Sheral HD, Lunday BJ (2013) On generating maximal nondominated Benders cuts. *Ann Oper Res* 210(1):57–72
- Stursberg P (2019) *On the mathematics of energy system optimization*. PhD thesis. Technische Universität München
- Vanderbeck F, Wolsey LA (2010) Reformulation and Decomposition of Integer Programs. In: Jünger M, Liebling TM, Naddef D, Nemhauser GL, Pulleyblank WR, Reinelt G, Rinaldi G, Wolsey LA (eds) *50 years of integer programming 1958–2008: from the early years to the state-of-the-art*. Springer, pp 431–502. isbn: 978-3-540-68279-0

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.