A NOVEL PARETO-OPTIMAL CUT SELECTION STRATEGY FOR BENDERS DECOMPOSITION

LUKAS GLOMB, FRAUKE LIERS, FLORIAN RÖSEL

All authors: FAU Erlangen-Nuremberg, Department Data Science, Cauerstraße 11, 91058 Erlangen, Germany

Abstract. Decomposition approaches can be used to generate practically efficient solution algorithms for a wide class of optimization problems. For instance, scenario-expanded two-stage stochastic optimization problems can be solved efficiently in practice using Benders Decomposition. The performance of the approach can be influenced by the choice of the cuts that are added during the course of the algorithm.

The so-called Magnanti-Wong method aims for Pareto-optimal cuts. Cuts based on minimal infeasible subsystems of a modified version of the sub problem have proven to provide strong cuts. Most recently, methods using facets of the sub problem's value function's epigraph have been developed.

We contribute to the field of cut selection strategies for Benders Decomposition by developing an innovative notion of Pareto-optimality, which implies an efficient cut selection strategy. The strategy aims for cuts that exclude a large set of points from being optimal. We further develop the algorithmic framework to exploit the potential of our cut selection strategy optimally. We benchmark our cut selection strategy against several other cut selection strategies on various instances. Some instances are taken from the MIPLib, some others are network design problems, and others are randomly generated mixed-integer linear programs. The computational results show that the developed method is, measured in CPU seconds needed to solve a problem to optimality, at least competitive for or better than the benchmark approaches for all three instance classes, especially, when it is combined as hybrid selection strategy with the minimal infeasible subsystem cut selection. In addition, the method clearly outperforms the other cut selection strategies measured in the number of cuts needed to solve a problem to optimality. Hence, the method is especially effective in situations with scarce memory or with a sub problem that is difficult to solve.

Keywords. Benders Decomposition, Cut Selection

1. Introduction and Literature Review

Originally published in Benders (1962) and developed for mixed-integer linear programs (**MILP** or **MIP**), the approach named Benders Decomposition (**BD**) works, on a meta-level, as follows. The variables of the optimization problem to be solved are separated into two classes, the so-called master variables and the sub variables. The approach alternately solves two optimization problems, the so-called master problem and the sub problem. In each iteration, a solution of the master problem is obtained. Fixing the variables from the master problem to the obtained values, a solution of the sub problem is determined and a new valid constraint is derived from the dual solution of the sub problem. This constraint is added to the master problem.

BD is the backbone of various solution methodologies for various optimization problems. It works exceptionally well, if the sub problem has a separable structure. This means that the sub problem

E-mail address: lukas.glomb@fau.de, frauke.liers@fau.de, florian.roesel@fau.de (CA's email).

represents a collection of smaller optimization problems that can be solved separately (but are coupled if the master variables are not fixed).

The high practical relevance of the approach is documented by a large number of articles. BD has been applied to solve a large number of often stochastic optimization problems with the property that fixing the first stage variable decomposes the problem into independent subproblems. In Baringo and Conejo (2011) the approach has been applied to maximize profits from wind power investments. This problem is highly affected by uncertainty, and hence the emerging scenario-expanded optimization model is decomposed into algorithmically tractable parts. BD can solve the otherwise intractable problem quickly. In Adulyasak et al. (2015), the approach is applied to solve a variant of the production routing problem under uncertainty. The scenario-expanded model, that is in practice difficult to solve due to its size, could be decomposed into smaller pieces, and the authors report the successful solution of instances of realistic size, and considerable speedups of BD in comparison to standard methods. The authors also report on several algorithmic improvements like lower-bound lifting inequalities, Pareto-optimal cuts (in the sense of Magnanti and Wong (1981)) and cut aggregation. They report that Pareto-optimal cuts have the most significant effect on CPU time (factor 2 to 4 on their instance set). In Maheo et al. (2019), BD is applied to solve a variant of the network design problem to create a public transit system for Canberra. The problem structure allows for splitting the problem into multiple independent sub problems after fixing only few variables. The authors used Pareto-optimal cuts in the sense of Magnanti and Wong (1981), and reported large speedups compared to standard solution methods. The authors Bärmann et al. (2015) could solve huge network design problems using an iterative aggregation procedure that dynamically expands subnetworks. It is shown that the expansion routines are closely related to feasibility cuts in the sense that the latter are implied by the expansion procedure. The method outperforms standard approaches by far when applied to railway network design problems. In Abdolmohammadi and Kazemi (2013), BD has been applied to optimize the utilization of combined heat and power systems. The authors report that the approach is superior to several benchmark methods. In Nasri et al. (2015), the approach has been applied to solve a network-constrained AC unit commitment problem with uncertain wind power production. It is used to decompose the emerging intractable scenario-expanded optimization problems into tractable parts, that can be solved independently. In Contreras et al. (2011), BD is very successfully applied to solve large-scale uncapacitated hub location problems. The problem as well decomposes into several sub problems if a small proportion of the variables is fixed. In Mansouri et al. (2020), the operational planning of energy hubs with demand response under uncertainty is optimized using BD, splitting the intractable scenario-expanded optimization model into tractable parts. The authors report considerable speedups in comparison to to standard methods. Grimm et al. (2017) have developed a generalized decomposition algorithm for a three-stage energy market problem that computes optimum electricity market price zones. They could show that welfare-maximum solutions can be computed within reasonable time although the problem itself is very complex. In Ambrosius et al. (2020) the approach is specified for the German electricity market, also taking network expansion as well as renewable energy into account. In Bayram and Yaman (2018), generalized BD has been applied to optimize the location of shelters and evacuation routes under uncertainty, being able to split the intractable scenario-expanded model into tractable parts. The authors applied several acceleration strategies, including Pareto-optimal cuts, and report significant CPU time savings. In You and Grossmann (2013), a multi-cut version of the approach is applied to solve a supply chain planning problem under uncertainty. The authors report considerable CPU time savings compared to the single-cut version.

Applications of BD to optimization problems without block structure exist as well. In Botton et al. (2013) it has been applied to efficiently handle extended formulations of the hop-constrained survivable network design problem. The authors report a considerable speedup compared to standard methods. In Qian et al. (2013), it has been successfully used to optimize communication networks. In Azad et al. (2013), it has been applied to optimize supply chain networks suffering facility and transportation disruptions. The authors report considerable CPU time savings, using the covering-cut-bundle method as presented in Saharidis et al. (2010), and maximum density cuts, compared to standard solution methods. In Fischetti et al. (2016), it is tuned to be applicable to capacitated facility location problems, while in Fischetti et al. (2017) it is applied to the uncapacitated version of this problem. The authors use several acceleration strategies and achieve considerable CPU time savings compared to alternative solution approaches for both problem versions. In Glomb et al. (2023b), a logic-based variant of BD has been applied to solve an integrated tail assignment and turnaround planning problem. The authors report considerable speedups compared to classical solution approaches. In Glomb et al. (2023c), it has been applied to solve the tail assignment problem suffering of part failure scenarios. Solving the huge scenario-expanded model with the decomposition led to significant CPU time savings in comparison to standard approaches.

Several approaches and techniques to accelerate BD are known. Many of them have been used in Rahmaniani et al. (2018), where the approach has been applied to solve large-scale stochastic network design problems. For this reason, this article is a good starting point to acquire knowledge about acceleration techniques. The authors of Geoffrion and Graves (1974) proposed a strategy that avoids solving the master problem to optimality in each iteration to attain performance benefits. In Cote and Laughton (1984) it has been shown that heuristic solutions of the master problems generate valid cuts. This has been specified, e.g., in Poojari and Beasley (2009), where a genetic algorithm is applied to generate solutions of the master problem. In McDaniel and Devine (1977) it is shown that valid cuts can also be calculated if relaxations of the master problems are used. Based on that, approaches to solve MIPs have been proposed that are often referred to as the two-phase-algorithm, that solve the linear programming relaxation of the master problem in the first phase, and reestablish the integrality constraints of the master problem in the second phase. The authors of Costa et al. (2012) proposed several strategies that aim for the determination of points in a superset of the master problem domain, for which it is beneficial to generate cuts. Several authors made the observation that a stabilization of the solution process of the master problems leads to performance gains. A concrete implementation of a stabilization procedure is Rei et al. (2009), where local branching has been used to limit the distance of subsequent master solutions. An alternative implementation of a stabilization approach has been set up in Santoso et al. (2005) by limiting the Hamming distance to a specified stabilization point. We want to note that our algorithmic framework makes use of modern MIP solvers being capable to insert additional constraints to an optimization problem on-the-fly during a single Branch-and-Cut run. The approach is called Branch-and-Benders-Cut, and is described, e.g., in Rahmaniani et al. (2017), and applied, e.g., in Gendron et al. (2016). The authors of Saharidis et al. (2011) point out that for several optimization problems it is beneficial to initialize the master problem with several valid inequalities, that can either be added upfront, or treated as cutting planes, that are added to the master problem during the solution process whenever it seems to be suitable. This strategy has also been applied in Fischetti et al. (2016).

An important step to improve the performance of BD is the cut selection strategy. The authors of Magnanti and Wong (1981) propose to generate a so-called non-dominated Benders cut, whenever the dual of the sub problem has multiple solutions. To achieve this, each generated cut is readjusted by solving an auxiliary version of the sub problem that depends on a so-called core point. Adaptations to this methods have been proposed. In Papadakos (2008), it is shown that clever updates of the core point accelerate the procedure further, and Sherali and Lunday (2013) shows that the solution of a perturbed version of the sub problem provides automatically a non-dominated cut, making the solution of a second optimization problem in each iteration unnecessary. For certain optimization problems, it seems to be beneficial to generate multiple cuts per master problem solution. In Saharidis et al. (2010), it is proposed to generate a so-called covering cut bundle, i.e., a set of Benders cuts that have as many master variables as possible in the union of their supports. In Fischetti et al. (2010) and in Stursberg (2019) cut selection strategies based on slightly adapted sub problems are proposed. We will present their contribution in Section 2 more precisely.

Some authors demonstrated the potential of making the approach more efficient by implementing problem specific techniques to strengthen the cuts, like, e.g., in Van Roy (1986) or Wentges (1996) for some variants of the well-known facility location problem. In Wu and Shahidehpour (2010), the approach is accelerated using several acceleration strategies like Pareto-optimal cuts or cuts with increased density, for network-constrained unit commitment problems. The authors report considerable speedups compared to a standard implementation of the approach.

In Bonami et al. (2020), several crucial numerical aspects regarding BD have been addressed. The authors provide a guideline how to implement the approach such that features of state-of-the-art MIP solvers, such as pre-solve, are actively exploited. The authors provide an extensive benchmark of their implementation and report considerable speedups.

Our contribution is the following. We develop a novel cut selection strategy, in the sense of Magnanti and Wong (1981), who presented Pareto-optimal Benders cuts. To implement this, we first develop a novel notion of Pareto-optimality. The idea is to compare the sets of points in the domain of the master problem, that have a chance to be the optimal solution of the overall problem after the corresponding cut is inserted into the master problem. Based on this, we develop a novel cut selection strategy, i.e., it generates, under mild assumptions, Pareto-optimal cuts according to our notion. The strategy only requires the solution of one linear program per generated cut, that is not considerably more difficult than the original sub problem. Further, we develop an algorithmic framework to apply this cut selection strategy. The framework specifies some details that are relevant for a good performance of the proposed strategy. We further conduct a computational study benchmarking our algorithmic framework against other cut selection strategies: the Magnanti-Wong method as presented in Magnanti and Wong (1981), cuts based on minimal infeasible subsystems (**MIS**) as presented in Fischetti et al. (2010), and facet defining cuts as presented in Stursberg (2019). We want to note that, even though we present straightforward motivation and proofs for our results, the method we propose amounts to a transfer of Conforti and Wolsey (2019) to the special case BD, following a suggestion made in the final remarks of Conforti and Wolsey (2019). Nevertheless, several algorithmic specifications we defined led to considerable performance benefits.

In Section 2 we introduce BD, as well as known cut selection strategies. Then, we propose our novel notion of Pareto-optimality, our cut selection strategy and prove several important statements about it. In Section 3 we present the algorithmic framework that enables to apply the cut selection strategy. In Section 4 we describe our computational study and its results. In Section (5) we summarize our results and give an outlook on future research directions.

2. A general Cut Selection Strategy for Benders Decomposition

In this section, we want to present a new approach to select valid cutting planes for BD. First, BD as originally presented in Benders (1962) is introduced. Afterwards, we will present and discuss known cut selection strategies. Some of these are variants of the popular Magnanti-Wong method originally presented in Magnanti and Wong (1981). Furthermore, we present a framework referred to as MIS-cuts, which has originally been presented in Fischetti et al. (2010). We also present an innovative cut selection strategy proposed by Stursberg (2019), that is capable of generating so-called facet generating cuts under mild assumptions. The main part of this section is the presentation of a new cut selection strategy that we developed based on the insights that we gained from the existing ones.

For a more detailed overview, we refer to Rahmaniani et al. (2017) which is an excellent survey about recent developments in this field.

2.1. **Benders Decomposition.** Originally presented in Benders (1962), the approach has been developed for optimization problems containing a set of "complicating", i.e., very often integer, variables, typically of the form

$$\min f^T x + c^T y \tag{1a}$$

$$\text{s.t. } Hx + Ay \geq b \tag{1b}$$

$$x \in X, y \geq 0, \tag{1c}$$

where $X \subset \mathbb{R}^n$ (the n-dimensional Euclidean space), $b \in \mathbb{R}^m$, $f \in \mathbb{R}^n$, $H \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^p$, $A \in \mathbb{R}^{m \times p}$.

The approach works straight-forwardly under the mild assumption that $\min_{x \in X} f^T x$ exists and a lower bound $\hat{\eta} \in \mathbb{R}$ for

$$\tilde{\eta} := \min_{x \in X} S(x),$$

where

$$S(x) := \min_{y \geq 0} \ c^T y \quad \text{s.t. } Ay \geq b - Hx,$$

exists and is easily obtainable. This is, e.g., the case, if c is non-negative. Then, the inner product $c^T y$ is certainly non-negative as well, and the choice $\hat{\eta} := 0$ is hence valid.

An artificial variable η along with the lower bound $\hat{\eta}$ is introduced, and afterwards the so-called master problem is solved repeatedly, which is defined as

$$\min \ f^T x + \eta \tag{2a}$$

$$\text{s.t. } 0 \geq \omega(x) \text{ for all } \omega \in \Omega^{\text{feas}} \tag{2b}$$

$$\eta \geq \omega(x) \text{ for all } \omega \in \Omega^{\text{opt}} \tag{2c}$$

$$\eta \geq \hat{\eta}, x \in X. \tag{2d}$$

Constraints (2b) and (2c) are the so-called Benders cuts and are frequently updated, as described in the following. The set $\Omega := \Omega^{\text{feas}} \cup \Omega^{\text{opt}}$ is initialized as $\emptyset$. Hence, in the first iteration, the solution $(\bar{x}, \bar{\eta})$ of Model (2) is generally $\bar{x} = \min_{x \in X} f^T x$, $\bar{\eta} = \hat{\eta}$.

After an optimal solution $(\bar{x}, \bar{\eta})$ of (2) is obtained, one solves the so-called dual sub problem, which is the linear programming dual of the primal sub problem,

$$\max \ \pi^T(b - H\bar{x}) \tag{3a}$$
$$\text{s.t. } \pi^T A \leq c^T \tag{3b}$$
$$\pi \geq 0. \tag{3c}$$

We note, that (3) is a linear program which is defined, independently on $\bar{x}$, on the polyhedron $P := \{\pi : \pi^T A \leq c^T, \pi \geq 0\}$. P is non-empty, implied by the assumption that $S(x)$ is bounded from below.

If the outcome of the optimization of (3) is "unbounded", we insert a so-called feasibility cut to Ω^{feas}, which reads as

$$\bar{\pi}^T(b - Hx) \leq 0,$$

where $\bar{\pi}$ is an extreme ray of P with a positive objective function value, i.e.,

$$\bar{\pi}^T A \leq 0, \ \bar{\pi}^T(b - H\bar{x}) > 0.$$

If the outcome of the optimization of (3) is "optimal", we insert a so-called optimality cut to Ω^{opt}, which reads as

$$\bar{\pi}^T(b - Hx) \leq \eta,$$

where $\bar{\pi}$ is an optimal solution of (3).

The procedure is repeated, until the optimal value of (3) coincides with $\bar{\eta}$. This happens after a finite number of steps, if extreme rays and optimal solutions of (3) are chosen out of a (necessarily finite) inner representation of P. For a detailed proof of convergence, we refer the interested reader to Benders (1962).

2.2. **Known Cut Selection Strategies.** If (3) has multiple optimal solutions in some iterations, the approach can be accelerated by choosing an appropriate one. In certain cases, the appropriate solution implying an efficient cut does not even have to be implied by an optimal solution of Model (3). In the following, some approaches defining which cuts should be used are described.

2.2.1. *The Magnanti-Wong Method.* A well-known cut selection strategy has been proposed by Magnanti and Wong (1981). The idea is to generate a non-dominated cut. A cut implied by a solution $\bar{\pi}$ of Model (3) dominates another cut implied by $\tilde{\pi}$, if and only if for all feasible master points $x \in X$

$$\bar{\pi}^T(b - Hx) \geq \tilde{\pi}^T(b - Hx),$$

with a strict inequality for at least one $x \in X$. We call a cut non-dominated, if no other cut dominates it. According to Magnanti and Wong (1981), a non-dominated cut can be obtained if a so-called core point, that is contained in the relative interior of the convex hull of the feasible master solutions, is known, i.e., a point $\hat{x} \in \text{ri}(\text{conv}(X))$. In this case, in each cut generating step Model (3) is solved first, providing a solution $\hat{\pi}$. Then the constraint

$$\pi^T(b - H\bar{x}) \geq \hat{\pi}^T(b - H\bar{x})$$

is inserted, while the objective function is changed to

$$\pi^T(b - H\hat{x}).$$

Then, the emerging Model is solved again to get a possibly new solution $\bar{\pi}$, which is guaranteed to define a non-dominated cut. For details regarding the proof of the statement, we refer to Magnanti and Wong (1981). The procedure has proven to accelerate the approach for various applications, e.g., Froyland et al.

(2014), to name only one. A drawback is the necessity to determine an appropriate core point, which is difficult for some problems. The authors of Papadakos (2008) suggest a strategy how an approximate core point can be determined on-the-fly with the help of an iterative procedure to overcome this drawback.

2.2.2. *Cuts defined by Minimal Infeasible Subsystems.* Fischetti et al. (2010) proposed a method to determine cuts using MIS of the slightly adapted variant the sub problem, which is the feasibility problem

$$\begin{aligned}\text{Find } & y \geq 0\\ \text{s.t. } & c^T y \leq \bar{\eta}\\ & Ay \leq b - H\bar{x}.\end{aligned}$$

Their computational study implies large speedups. Practically, the approach is implemented by adding a variable and a constraint to Model (3). The modified dual sub problem then reads

$$\max\ \pi^T(b - H\bar{x}) - \pi_0\bar{\eta} \tag{4a}$$

$$\text{s.t. } \pi^T A \leq \pi_0 c^T \tag{4b}$$

$$\pi^T w + \pi_0 w_0 = 1 \tag{4c}$$

$$\pi \geq 0. \tag{4d}$$

The program uses an appropriately dimensioned positive-valued weight vector w, which can be arbitrarily chosen. The authors of Fischetti et al. (2010) suggest to choose w_i to be 1 for $i = 0$ and for indices i for which the corresponding row of H is not the zero row, and 0 otherwise.

The cut that is generated by a solution $(\bar{\pi}, \bar{\pi}_0)$ of the modified dual sub problem (4) would read

$$\bar{\pi}^T(b - Hx) \leq \bar{\pi}_0 \eta$$

An advantage of the proposed framework is that the feasibility cuts and optimality cuts can be derived in the same way, as dual solutions instead of extreme rays, since we note that in case $\bar{\pi}_0$ is 0, we get a cut that does not depend on the variable η. Hence, in this case it is a feasibility cut.

Furthermore, the authors of Fischetti et al. (2010) report large computational benefits compared to alternative implementations.

Even though the MIS approach is performing extraordinarily well in practice, it has several drawbacks. It might generate cuts that do not support the epigraph of the sub problem's value function, as defined later in (5), in any point. As a result, multiple cuts might have to be generated in order to cut off a single assignment of master variables. Furthermore, the scaling of the constraints influences the behavior of the MIS approach, as well as the existence of constraints that are unnecessary in a sense that they are implied by other constraints. Furthermore, the strategy cannot be applied straight-forwardly to solve problems with equality constraints, leading to dual variables that are not restricted in sign. Several adaptations to the framework have to be made in this case. Hence, especially if the optimization problem to be solved is formulated rather poorly, MIS cut selection might not be the means of choice.

2.2.3. *Facet Generating Cuts.* Most recently, Stursberg (2019) proposed a cut selection framework which has some parallels to MIS cut selection. The relevant results of the dissertation have also been published in the journal article Brandenberg and Stursberg (2021). The idea is to aim for cuts that represent facets of the polyhedral epigraph of the sub problem's value function,

$$\mathrm{E} := \{(x, \eta) \mid \eta \geq \min c^T y\colon Ay \geq b - Hx, y \geq 0\}. \tag{5}$$

The theory developed in Stursberg (2019) is based on reverse polar sets of the epigraph. Since immediate representations of reverse polar sets are likely to be intractable, the author solves optimization problems defined over the so-called Relaxed Alternative Polyhedron, that are under mild assumptions equivalent to the optimization problems over the reverse polar set. The Relaxed Alternative Polyhedron is defined as

$$P^{\leq}(x,\eta) := \{\pi, \pi_0 \geq 0 \mid \pi^T A + \pi_0 c^T = 0,\ \pi^T(b - Hx) + \pi_0 \eta \geq 1\}. \tag{6}$$

According to Theorem 3.32 of Stursberg (2019), there is an optimal extremal point of the Relaxed Alternative Polyhedron inducing a cut that supports the epigraph in one of its facets under mild assumptions on the objective function, or, as the author calls it, weight vector, that we denote as $(\tilde{w}, \tilde{w}_0)$. The author states further, that the dimension of the set of weight vectors leading necessarily to a facet defining cut is higher than the dimension of the set of weight vectors that might lead to a cut that is not facet defining. Further, a procedure is proposed which deterministically leads to a facet defining cut, even though the author claims that its application is computationally inefficient in many cases.

In the empirical study, it turns out that setting $(\tilde{w}, \tilde{w}_0)$ to $(H(\tilde{x} - \bar{x}), \tilde{\eta} - \bar{\eta})$ is quite efficient. The vector $(\tilde{x}, \tilde{\eta})$ is a, possibly frequently updated, feasible solution of the master problem and an upper bound on the corresponding sub problem objective value, that has been obtained heuristically beforehand. The vector $(\bar{x}, \bar{\eta})$ is the current master solution and the corresponding sub problem objective estimate. The author calls this strategy “adaptive cuts”. Since this seems to be an efficient choice of objective functions, we will use this strategy throughout in our benchmarks, calling the cuts “facet generating” or simply “facet”.

The author proposes furthermore several equivalent variants of sub problems than can be chosen when the “facet” cut selection is applied. One is the following, slightly adapted from Stursberg (2019).

$$\max\ \pi^T(b - H\bar{x}) - \pi_0\bar{\eta} \tag{7a}$$

$$\text{s.t. } \pi^T A \leq \pi_0 c^T \tag{7b}$$

$$\pi^T H(\tilde{x} - \bar{x}) + \pi_0(\tilde{\eta} - \bar{\eta}) = 1 \tag{7c}$$

$$\pi \geq 0. \tag{7d}$$

We note that the difference of the MIS sub problem and the facet generating sub problem reduces to an adapted normalization constraint: While (4c) is the normalization constraint of Model (4), Constraint (7c) is the normalization constraint of Model (7). Apart from the normalization constraint, the two models are identical.

Remark. While MIS cut selection relies on sign-restricted dual variables, the facet approach can also be applied when the dual variables are free. Furthermore, as discussed before, MIS cut selection depends, e.g., on the scaling of certain constraints. In contrast to that, the author of Stursberg (2019) claims that a cut that is generated using Model (7) supports the epigraph of the sub problem value function at a point on the connection line of the iterate $(\bar{x}, \bar{\eta})$ and the incumbent $(\tilde{x}, \tilde{\eta})$. This implies especially, that the facet technique is independent from algebraic transformations of the optimization problems under consideration, that leave the epigraph of the sub problem unaffected.

2.3. **β-Dominance and our Cut Selection Strategy.** In the following, we propose a cut selection strategy with the goal to generate cuts which are suited to solve the overall optimization problem (1) efficiently. The idea of the new strategy incorporates knowledge about the master problem and solutions obtained in previous iterations of the algorithm.

2.3.1. *An Introductory Example.* We first motivate the strategy we propose. The idea of the strategy is that in some cases it is not beneficial to solve the dual sub problem to optimality. Instead, it is beneficial to determine a dual solution that leads to a cut that excludes as many points as possible (including the current master iterate) from having a chance to be the optimal solution of the overall optimization problem. The fact that an optimal dual solution does not necessarily lead to such a cut is demonstrated with the following example.

Example 2.1. *Consider the following optimization problem, which is defined on the Euclidean plane:*

$$\min -\frac{1}{10}x + y \tag{8a}$$

$$s.t.\ \ y \geq \frac{1}{4}x \tag{8b}$$

$$y \geq -x \tag{8c}$$

$$y \geq -1 - 2x \tag{8d}$$

$$x \in [-2, 2]. \tag{8e}$$

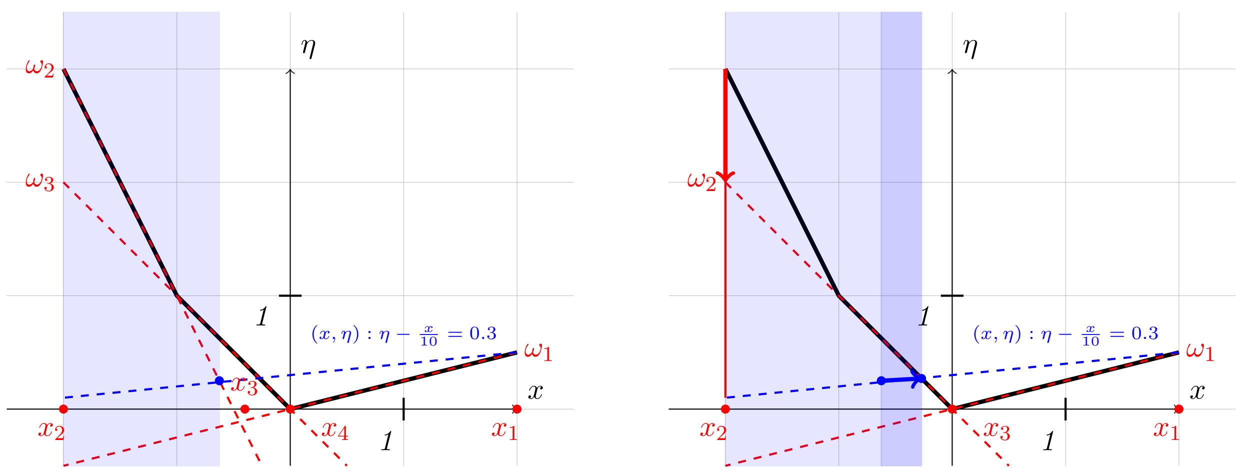

FIGURE 1. Value function of sub problem, and iterates of BD: Left describes the classical course where optimal solutions of the dual sub problem are used, Right describes the course where β-dominant cuts are used. The number of iterations required to solve the problem is reduced.

Since $\bar{x} = 2$ *optimizes* $\frac{-1}{10}x$ *on the interval* $[-2, 2]$, *we initialize the set of cuts* $\Omega = \{\omega_1\}$ *with* $\omega_1(x) = \frac{1}{4}x$, *and set up the master problem as*

$$\begin{aligned} &\min -\frac{1}{10}x + \eta \\ &s.t.\ \ \eta \geq \omega(x) \quad \textit{for all } \omega \in \Omega. \end{aligned}$$

New cuts, as well as the already calculated cut ω_1, are obtained from the dual sub problem:

$$\max \pi_1 \cdot \frac{1}{4}x + \pi_2 \cdot (-x) + \pi_3 \cdot (-1 - 2x) \tag{9a}$$

$$\pi_1 + \pi_2 + \pi_3 = 1 \tag{9b}$$

$$\pi \geq 0. \tag{9c}$$

The sub problem in this form just evaluates the maximum of the three functions ($\frac{1}{4}x$), ($-x$), and ($-1-2x$), with the freedom to generate a convex combination of all maxima, if the maximum is not unique.

The course of the classical version (using dual ***optimal*** *solutions) of BD is shown in the left side of Figure 1. The algorithm first finds the point which minimizes the pure master objective, which is $\bar{x} = 2$. Then, it solves the sub problem to generate the cut $\eta \geq \frac{1}{4}x$. The next point evaluates to $(\bar{x}, \bar{\eta}) = (-2, -0.5)$ with an objective value of -0.3. Then, the sub problem is solved again to generate the cut $\eta \geq -1 - 2x$. The next point evaluates to $(\bar{x}, \bar{\eta}) = (\frac{-4}{9}, \frac{-1}{9})$ with an objective value of $\frac{-1}{15}$. Then, the sub problem is solved again to generate the cut $\eta \geq -x$. The next point evaluates to $(\bar{x}, \bar{\eta}) = (0, 0)$, with an objective value of 0. This is the optimal point, since $\bar{\eta}$ equals the optimal value of the sub problem at this point.*

We investigate a set which is defined in terms of the best solution to the master problem found in the first iteration of the algorithm, which is $(\tilde{x}, \tilde{\eta}) = (2, 0.5)$, with a combined value $\beta = 0.3$. The set contains all points which are candidates to be optimal solutions of the overall optimization problem, i.e.,

$$\{x \in X \mid \exists \eta \colon f^T x + \eta \leq f^T \tilde{x} + \tilde{\eta} \text{ and } \eta \geq \omega(x) \text{ for all } \omega \in \Omega\}.$$

After the first iteration, this set reads

$$\{x \in [-2, 2] \mid \exists \eta \colon \frac{-1}{10}x + \eta \leq 0.3 \text{ and } \eta \geq \frac{1}{4}x\} = [-2, 2].$$

In the next iteration, the algorithm chooses exactly one candidate $\bar{x}$ out of this set. The sub problem is solved to optimality, and a cut is derived that especially makes sure that for this candidate all feasible assignments of η in future iterations of the algorithm have to be greater or equal to the optimal value of the sub problem attained for this candidate $\bar{x}$.

This especially makes sure, that no non-optimal point in X is visited more than once in the course of the algorithm. Nevertheless, this can also be assured, if we postulate that a cut leads to a new set of candidates to be optimal solutions, which excludes the current iterate. This is always the case, if we choose a dual solution $\bar{\pi}$ of the sub problem, that fulfills

$$\bar{\pi}(b - H\bar{x}) > f^T \tilde{x} - f^T \bar{x} + \tilde{\eta}.$$

Starting again from the beginning of iteration two of the prior run of the algorithm, this extends the possibilities to choose a valid cut, cutting off point $\bar{x} = -2$. E.g., one could choose $\pi_2 = 1$, leading to the cut $\eta \geq -x$, or the choice $\pi_1 = 0.5$, $\pi_3 = 0.5$, leading to the cut $\eta \geq -0.5 - \frac{7}{8}x$.

Depending on what cut is chosen, the set of candidate optimal solution in the next iteration varies. The original cut $\eta \geq -1 - 2x$ produces as candidate set the interval $[\frac{-13}{21}, 2]$, while the cut $\eta \geq -x$ produces $[\frac{-3}{11}, 2]$, which is a strict subset. This fact reflects also in the course of the standard version, which chooses $x = \frac{-4}{9}$ as next candidate solution, which is not in the candidate set of the cut $\eta \geq -x$, and hence would have been excluded upfront if this cut would have been chosen.

Overall, the choice of $\eta \geq -1 - 2x$ as second cut leads to three cuts in total, while choice of $\eta \geq -x$ leads to two cuts in total.

Nevertheless, we note that the cut $\eta \geq -1-2x$ is non-dominated in the notion of Magnanti and Wong (1981), and furthermore also corresponds to a MIS of the sub problem, since the dual solution corresponds to a single violated constraint. It defines a facet of the epigraph as well.

The example demonstrates that the existing cut selection strategies might miss to choose the clearly superior cut. Hence, we next present a new technique for generating Benders cuts.

2.3.2. *Cut Selection Paradigm.* The cuts in the example aimed for minimizing the size of the set of points in the master problem domain that are potentially better than an already obtained solution. This is the paradigm our cut selection strategy follows. Hence, we formally define a solution candidate set that depends on the value of a feasible solution $\tilde{x} \in X$ and the corresponding sub problem optimal value $\tilde{\eta} \in \mathbb{R}$. This value is called **budget** throughout and is denoted as β. It holds that $\beta := f^T\tilde{x} + \tilde{\eta}$.

Definition 2.2 (Solution Candidate Set.)**.** *Given the master problem* (2) *with feasibility and optimality cuts $\Omega = \Omega^{feas} \cup \Omega^{opt}$, and $\beta \in \mathbb{R}$, we define the* solution candidate set (**SC-set**) *as*

$$C_{\Omega,\beta} := \{x \in X \mid \exists \eta \colon f^T x + \eta \leq \beta, 0 \geq \omega(x) \text{ for } \omega \in \Omega^{feas}, \eta \geq \omega(x) \text{ for } \omega \in \Omega^{opt}\}.$$

For a feasible solution $\tilde{x}$ of Model (2) *with $\tilde{\eta}$ denoting the corresponding sub problem optimal value, we define the* SC-set *analogously as*

$$C_{\Omega,\tilde{x},\tilde{\eta}} := C_{\Omega,f^T\tilde{x}+\tilde{\eta}}.$$

We first observe, that the SC-set of a system of cuts and for a budget which is the value of the best solution found so far is exactly the set where a potential optimal solution of (1) can be.

Observation 2.3. *Given Problem* (1)*, and a system of valid cuts Ω, it holds that the set of points in X that can be completed to an optimal solution of* (1) *is a subset of $C_{\Omega,\beta}$ for all budgets β that are not lower than the optimal value of* (1)*.*

We can state a criterion that excludes a certain point in X from the SC-set for a certain budget.

Observation 2.4. *Given a feasible master solution $\bar{x} \in X$ and a budget $\beta \in \mathbb{R}$. Given an optimality cut $\omega_{\bar{\pi}}$ induced by a feasible solution $\bar{\pi}$ of* (3)*, it holds that*

$$\bar{x} \in C_{\omega_{\bar{\pi}},\beta} \Leftrightarrow f^T\bar{x} + \bar{\pi}^T(b - H\bar{x}) \leq \beta.$$

The observation implies, that a version of the approach, that generates a cut that has a value of greater than the current budget at the point where it is generated has the property that the algorithm never visits a point in X twice before it terminates.

We further note that an algorithm that follows that paradigm also has the possibility to exclude points in X with an infeasible sub problem by optimality cuts instead of feasibility cuts.

Next, we propose a dominance criterion similar to that of Magnanti and Wong (1981), which is based on SC-sets. One obvious advantage is that it includes also feasibility cuts, in contrast to the dominance notion of Magnanti and Wong (1981).

Definition 2.5 (β-Dominance)**.** *Given the master problem* (2) *with valid cuts ω_1 and ω_2, cut ω_1 β-dominates cut ω_2 for a budget $\beta \in \mathbb{R}$, if*

$$C_{\omega_1,\beta} \subsetneq C_{\omega_2,\beta}.$$

In the following we want to point out the connection between β-dominance and the dominance notion of Magnanti and Wong (1981).

Theorem 2.6. *Given the master problem* (2) *with a convex feasible set X and two optimality cuts ω_1 and ω_2 with*

$$\omega_1(x) \geq \omega_2(x) \text{ for all } x \in X \text{ and } \omega_1(x) > \omega_2(x) \text{ for at least one } x \in X.$$

Then, ω_1 β-dominates ω_2 for all budgets β in

$$\mathcal{B} := (\inf_{\tilde{\beta}} C_{\omega_2,\tilde{\beta}} \cap X \neq \emptyset, \sup_{\tilde{\beta}} C_{\omega_1,\tilde{\beta}} \cap X \neq X).$$

Proof. Let $\bar{x} \in C_{\omega_1,\beta}$ for an arbitrary β. Then, there is an $\eta \in \mathbb{R}$ such that

$$f^T\bar{x} + \eta \leq \beta \text{ and } \eta \geq \omega_1(\bar{x}).$$

Since $\omega_1(\bar{x}) \geq \omega_2(\bar{x})$, this implies that for this choice of η

$$f^T\bar{x} + \eta \leq \beta \text{ and } \eta \geq \omega_2(\bar{x})$$

also applies, and hence $\bar{x} \in C_{\omega_2,\beta}$. Assume, that there is a budget β in $\mathcal{B}$ such that $C_{\omega_1,\beta} = C_{\omega_2,\beta}$. Let $\bar{x}$ be a point for which $\omega_1(\bar{x}) > \omega_2(\bar{x})$. If $\bar{x} \in C_{\omega_1,\beta}$, we choose $\tilde{x} \notin C_{\omega_1,\beta}$, if $\bar{x} \notin C_{\omega_1,\beta}$, we choose $\tilde{x} \in C_{\omega_1,\beta}$. Both choices are guaranteed by the properties of $\mathcal{B}$.

We consider the case with $\bar{x} \in C_{\omega_1,\beta}$ first. This implies, that

$$f^T\bar{x} + \omega_2(\bar{x}) < f^T\bar{x} + \omega_1(\bar{x}) \leq \beta.$$

On the other hand, we have

$$f^T\tilde{x} + \omega_1(\tilde{x}) \geq f^T\tilde{x} + \omega_2(\tilde{x}) > \beta.$$

We consider for $i \in \{1, 2\}$ the functions

$$g_i \colon \lambda \mapsto f^T(\bar{x} + \lambda(\tilde{x} - \bar{x})) + \omega_i(\bar{x} + \lambda(\tilde{x} - \bar{x}))$$

Since ω_i are affine functions, g_i is also an affine function. We conclude that since $g_i(0) \leq \beta$ and $g_i(1) \geq \beta$ with one strict inequality each, there is exactly one point in $[0, 1]$ where g_i evaluates to β. Let λ_i be the point in $[0, 1]$ with $g_i(\lambda_i) = 0$. Since $g_2(0) < g_1(0)$ and $g_2(1) \leq g_1(1)$, $\lambda_2 > \lambda_1$ holds. If we choose $\bar{\lambda} = \frac{\lambda_1+\lambda_2}{2}$, then $\bar{x} + \bar{\lambda}(\tilde{x} - \bar{x})$ is a point in X which is contained in $C_{\omega_2,\beta}$, but not in $C_{\omega_1,\beta}$. This is a contradiction to our assumption. The second case is analogue. This proves the statement. ☐

We note that the set $\mathcal{B}$ denotes the budgets, for which at least one of the cuts ω_1, ω_2 has a non-trivial, i.e., $\notin \{\emptyset, X\}$, candidate set. Theorem 2.6 has the consequence, that if a cut is generated which is not β-dominated for reasonable budgets, it is also non-dominated in the sense of Magnanti and Wong (1981) under mild assumptions. The opposite direction is not true, i.e., it is possible that cut that is non-dominated in the sense of Magnanti and Wong (1981) is β-dominated, as already demonstrated in Example 2.1.

The next theorem states, that a valid cut that β-dominates another cut does this also for smaller budgets under certain conditions. This gives further justification for choosing the dominating cut over the dominated one, if these conditions apply for two cuts.

Theorem 2.7. *Given the master problem* (2) *with a convex feasible set X, a point $\bar{x} \in X$, a budget $\bar{\beta}$ and two optimality cuts ω_1 and ω_2 with*

$$C_{\omega_1,\bar{\beta}} \subseteq C_{\omega_2,\bar{\beta}} \text{ and } \omega_2(\bar{x}) > \omega_1(\bar{x}) \geq \bar{\beta} - f^T\bar{x}.$$

Then it holds that

$$C_{\omega_1,\beta} \subsetneq C_{\omega_2,\beta}$$

for all $\beta \in \mathcal{B}$, *where*

$$\mathcal{B} := (\inf_{\tilde{\beta}} C_{\omega_1,\tilde{\beta}} \neq \emptyset, \bar{\beta}).$$

Proof. Assume there is a $\beta \in \mathcal{B}$ for which

$$C_{\omega_1,\beta} \nsubseteq C_{\omega_2,\beta}, \text{ i.e., } \exists \tilde{x} \in X \text{ with } \omega_1(\tilde{x}) + f^T\tilde{x} \leq \beta,\ \omega_2(\tilde{x}) + f^T\tilde{x} > \beta.$$

We first note that $\omega_1(\bar{x}) + f^T\bar{x} > \bar{\beta}$, since otherwise $\bar{x} \in C_{\omega_1,\bar{\beta}}$ and $\bar{x} \notin C_{\omega_2,\bar{\beta}}$, which is a contradiction to our assumption. This implies that $\tilde{x} \neq \bar{x}$. We further note that $\omega_2(\tilde{x}) + f^T\tilde{x} \leq \bar{\beta}$ since otherwise $\tilde{x} \in C_{\omega_1,\bar{\beta}}$ and $\tilde{x} \notin C_{\omega_2,\bar{\beta}}$, which is a contradiction to our assumption. Especially, this implies that $\bar{\beta} < f^T\bar{x} + \omega_1(\bar{x}) < f^T\bar{x} + \omega_2(\bar{x})$ and $f^T\tilde{x} + \omega_1(\tilde{x}) < f^T\tilde{x} + \omega_2(\tilde{x}) \leq \bar{\beta}$.

We consider for $i \in \{1, 2\}$ the affine functions

$$g_i \colon \lambda \mapsto f^T(\bar{x} + \lambda(\tilde{x} - \bar{x})) + \omega_i(\bar{x} + \lambda(\tilde{x} - \bar{x})).$$

We note that $g_2(\lambda) > g_1(\lambda)$ for all $\lambda \in [0, 1]$. We further note that $g_1(0) > \bar{\beta}$ and $g_1(1) < \bar{\beta}$. Since g_1 is continuous, there exists $\bar{\lambda} \in [0, 1]$ with $g_1(\bar{\lambda}) = \bar{\beta}$. Since g_2 is strictly larger than g_1 on $[0, 1]$, we have that $g_2(\bar{\lambda}) > \bar{\beta}$. Hence, $\bar{x} + \bar{\lambda}(\tilde{x} - \bar{x})$ is contained in $C_{\omega_1,\bar{\beta}}$, but not in $C_{\omega_2,\bar{\beta}}$. This is a contradiction to our assumption and we can conclude that $C_{\omega_1,\beta} \subseteq C_{\omega_2,\beta}$. Assume there is a $\beta \in \mathcal{B}$ for which

$$C_{\omega_1,\beta} = C_{\omega_2,\beta}.$$

Then there exists $\tilde{x} \in X$ such that

$$\omega_1(\tilde{x}) + f^T\tilde{x} = \beta,$$

since $\emptyset \neq X \setminus C_{\omega_1,\beta}$ and $\emptyset \neq C_{\omega_1,\beta}$. It further holds that

$$\omega_2(\tilde{x}) + f^T\tilde{x} = \beta$$

as well, since $\omega_2(\tilde{x}) + f^T\tilde{x} > \beta$ would imply that the candidate sets are not equal, and if $\omega_2(\tilde{x}) + f^T\tilde{x} < \beta$ would hold we could define $g_i, i = 1, 2$ similar as before, and could observe that there is a small $\varepsilon > 0$ such that $g_1(\varepsilon) > \beta$, but $g_2(\varepsilon) < \beta$, what would also imply that the candidate sets are not equal.

So, to conclude our argument, we take $g_i, i = 1, 2$ as defined before. We note that $g_2(\lambda) > g_1(\lambda)$ for $\lambda \in [0, 1)$. Since g_1 is continuous, this interval contains a point $\bar{\lambda}$ for which $g_1(\bar{\lambda}) = \bar{\beta}$, and this implies that $g_2(\bar{\lambda}) > \bar{\beta}$. This implies that $\bar{x} + \bar{\lambda}(\tilde{x} - \bar{x}) \in C_{\omega_1,\bar{\beta}}$, but not in $C_{\omega_2,\bar{\beta}}$. This is a contradiction to our assumption and concludes the proof. □

2.3.3. *The role of feasibility cuts.* The focus of the presented theory is on optimality cuts. This is appropriate, since the theory proposes that each feasible solution of the dual sub problem with a value high enough defines a valid cut that cuts off the current master solution.

The question arises if there is a need for feasibility cuts at all. We discuss in the following, in which cases which kind of cut should be added.

The SC-set of a feasibility cut $\omega_{\bar{\pi}}$, induced by an extreme ray $\bar{\pi}$ of (3), in general does not depend on the budget β:

$$C_{\omega_{\bar{\pi}},\beta} = \{x \in X \colon \omega_{\bar{\pi}}(x) \leq 0\}$$

The SC-set of an optimality cut gets smaller with β, while the SC-set of a feasibility cut remains constant. From that perspective, optimality cuts are preferable. Nevertheless, it is possible that a feasibility cut is the only non-dominated valid cut for some budgets β.

In the following, we describe how cuts can be obtained that are not β-dominated.

2.3.4. *Optimal line shifts.* In the following we want to introduce an approach to generate Benders cuts, that are, under certain conditions, non-dominated in the sense of Definition 2.5.

The idea is to exclude as many points on a line from further consideration. We call this line shift.

Definition 2.8 (Line shift.). *Given the master problem* (2) *with a feasible set X, and points $\tilde{x} \in X$, $\bar{x} \in X$, and $\beta \in \mathbb{R}$ with the property that*

$$f^T\tilde{x} + S(\tilde{x}) \leq \beta \quad \textit{and } f^T\bar{x} + S(\bar{x}) \geq \beta,$$

we define a line shift *induced by $(\bar{x}, \tilde{x}, \beta)$ as an optimality cut $\bar{\omega}$, for which a parameter $\mu \in [0, 1]$ exists, such that*

$$\begin{aligned} f^T(\bar{x} + \mu(\tilde{x} - \bar{x})) + \bar{\omega}(\bar{x} + \mu(\tilde{x} - \bar{x})) &= \beta, \\ f^T\bar{x} + \bar{\omega}(\bar{x}) &\geq \beta, \\ f^T(\bar{x} + \tilde{\mu}(\tilde{x} - \bar{x})) + \tilde{\omega}(\bar{x} + \tilde{\mu}(\tilde{x} - \bar{x})) &< \beta \textit{ for all } \tilde{\mu} \in (\mu, 1), \end{aligned}$$

or a feasibility cut, $\bar{\omega}$, for which a parameter $\mu \in [0, 1]$ exists, such that

$$\begin{aligned} \bar{\omega}(\bar{x} + \mu(\tilde{x} - \bar{x})) &= 0, \\ \bar{\omega}(\bar{x}) &> 0, \\ \bar{\omega}(\bar{x} + \tilde{\mu}(\tilde{x} - \bar{x})) &< 0 \textit{ for all } \tilde{\mu} \in (\mu, 1). \end{aligned}$$

In both cases, we call μ the depth *of the line shift.*

We observe that for all points $\bar{x} \in X$, arbitrary feasibility cuts or optimality cuts $\bar{\omega}$ with $f^T\bar{x}+\bar{\omega}(\bar{x}) \geq \beta$ are line shifts with a certain depth. The following theorem states that under certain conditions determining a line shift with a depth that is as high as possible provides a cut that is not β-dominated. We call a line shift with maximum depth optimal line shift (**OLS**).

Theorem 2.9. *Given the master problem* (2) *with a convex feasible set X, and points $\tilde{x} \in X$, $\bar{x} \in X$, and $\beta \in \mathbb{R}$. A line shift $\bar{\omega}$ induced by $(\bar{x}, \tilde{x}, \beta)$ with maximal depth $\bar{\mu} < 1$ is not β-dominated, if*

$$\hat{x} := \bar{x} + \bar{\mu}(\tilde{x} - \bar{x}) \in \textit{relint}(X).$$

Proof. We provide the proof for the statement that a line shift that is an optimality cut is not β-dominated by another optimality cut. This proof can be easily adapted for the cases with feasibility cuts. For this reason, we omit the analogue reasoning.

Assume there is a valid Benders cut $\tilde{\omega}$ that β-dominates $\bar{\omega}$, i.e.,

$$C_{\tilde{\omega},\beta} \subsetneq C_{\bar{\omega},\beta} \Leftrightarrow C_{\tilde{\omega},\beta} \setminus C_{\bar{\omega},\beta} = \emptyset.$$

We first note that $\bar{\mu} < 1$ implies that $f^T\bar{x} + \bar{\omega}(\bar{x}) > \beta$. This in turn implies that $f^T\bar{x} + \tilde{\omega}(\bar{x}) > \beta$ holds, since otherwise $\bar{x} \in C_{\tilde{\omega},\beta}$, but $\bar{x} \notin C_{\bar{\omega},\beta}$. We further note that the depth of $\tilde{\omega}$ is $\bar{\mu}$ as well, since if the depth of $\tilde{\omega}$ would be $\tilde{\mu} < \bar{\mu}$, then $\hat{x} \in C_{\tilde{\omega},\beta}$, but $\hat{x} \notin C_{\bar{\omega},\beta}$.

Our assumption implies that a point $\breve{x} \in X$ exists with $f^T\breve{x} + \tilde{\omega}(\breve{x}) > \beta$, but $f^T\breve{x} + \bar{\omega}(\breve{x}) \leq \beta$. Since $\hat{x} \in \text{relint}(X)$, there exists a parameter $\theta > 1$, such that $\breve{x} + \theta(\hat{x} - \breve{x}) \in X$. This in turn implies that $f^T\breve{x} + \bar{\omega}(\breve{x}) = \beta$, as well as $f^T(\breve{x} + \theta(\hat{x} - \breve{x})) + \bar{\omega}(\breve{x} + \theta(\hat{x} - \breve{x})) = \beta$.

We note that $f^T x + \tilde{\omega}(x) > \beta$ has to hold for all x in a relatively open neighborhood of $\frac{\breve{x}+\hat{x}}{2}$. Using the same argument as before, on this open neighborhood it holds that $f^T(x) + \bar{\omega}(x) = \beta$. Hence, $f^T x + \bar{\omega}(x)$

is constant on X. This implies that $\bar{\omega}$ is a line shift with depth 1, which contradicts the assumptions made upfront. $\square$

As a consequence, we would be able to determine cuts that are non-dominated in the sense of Definition 2.5, if we could determine OLS. The following observation states that an OLS can be obtained by solving a certain optimization problem.

Observation 2.10. *Given Problem* (1) *and the corresponding master problem* (2) *with feasible set* X*, a budget* $\beta \in \mathbb{R}$ *and points* $\tilde{x} \in X$*,* $\bar{x} \in X$*, with* $f^T\bar{x} + S(\bar{x}) \geq \beta$*,* $f^T\tilde{x} + S(\tilde{x}) \leq \beta$*. If no optimality cut* $\bar{\omega}$ *exists for which* $\bar{\omega}(\tilde{x}) + f^T\tilde{x} = \beta$ *and* $\bar{\omega}(\bar{x}) + f^T\bar{x} = \beta$*, then a solution* $(\bar{\pi}, \bar{\mu})$ *of the following optimization problem defines an optimality cut that is a line shift induced by* $(\bar{x}, \tilde{x}, \beta)$ *with depth* $\bar{\mu} \leq 1$*, and no optimality cut that is a line shift induced by this triple has a higher depth.*

$$\max \quad \mu \tag{10a}$$

$$s.t.\ \pi^T(b - H(\bar{x} + \mu(\tilde{x} - \bar{x}))) + f^T(\bar{x} + \mu(\tilde{x} - \bar{x})) = \beta \tag{10b}$$

$$\pi^T A \leq c^T \tag{10c}$$

$$\pi^T(b - H\bar{x}) + f^T\bar{x} \geq \beta \tag{10d}$$

$$\pi \geq 0,\ \ \mu \geq 0. \tag{10e}$$

Especially, the optimization problem has at least one feasible solution, and it is not unbounded.

Containing products of variables, Model (10) is not a linear program anymore, since it contains the bilinear constraint (10b). Nevertheless, Model (10) can be reformulated to a fractional linear program, and under some mild assumptions which are easy to check on the fly of an algorithmic BD approach, Model (10) can be reformulated, following the method presented in Charnes and Cooper (1962), to an equivalent linear program with one variable that is restricted to be strictly positive. This is the statement of Lemma 2.11.

Lemma 2.11. *Given Problem* (1) *and the corresponding master problem* (2) *with feasible set* X*, a budget* $\beta \in \mathbb{R}$ *and points* $\tilde{x} \in X$*,* $\bar{x} \in X$*, with* $f^T\bar{x} + S(\bar{x}) \geq \beta \geq f^T\tilde{x} + S(\tilde{x})$*.*

i) If further no optimality cut $\bar{\omega}$ *exists for which* $\bar{\omega}(\tilde{x}) + f^T\tilde{x} = \beta$ *and* $\bar{\omega}(\bar{x}) + f^T\bar{x} = \beta$*, then, each optimal solution* $(\bar{\rho}, \bar{\alpha})$ *of*

$$\max \quad \rho^T(b - H\bar{x}) + \alpha(f^T\bar{x} - \beta) \tag{11a}$$

$$\rho^T H(\tilde{x} - \bar{x}) - \alpha f^T(\tilde{x} - \bar{x}) = 1 \tag{11b}$$

$$\rho^T A \leq \alpha c^T \tag{11c}$$

$$\alpha \geq 0,\ \ \rho \geq 0. \tag{11d}$$

has the property, that $\bar{\pi} := \frac{\bar{\rho}}{\bar{\alpha}}$*,* $\bar{\mu} := \bar{\rho}^T(b - H\bar{x}) + \bar{\alpha}(f^T\bar{x} - \beta)$ *is an optimal solution of Model* (10)*, if* $\bar{\alpha} > 0$*.*

ii) If $S(\bar{x}) < \infty$*, then* $\bar{\alpha} > 0$*.*

Remark. As a consequence of Lemma 2.13 and Theorem 2.15 that are presented later in this paper, it follows that the condition $S(\bar{x}) < \infty$ is not restrictive. Even if it does not hold, Model (11) can be used as surrogate for the dual sub problem.

Proof. For i), it is easy to check, that Constraint (10b) is equivalent to

$$\mu = \frac{\pi^T(b - H\bar{x}) + f^T\bar{x} - \beta}{\pi^T H(\tilde{x} - \bar{x}) - f^T(\tilde{x} - \bar{x})},$$

if we can prove that the denominator is nonzero over the feasible set of Model (10). Hence, μ can be replaced in the objective function of Model (10). The result is an optimization problem with linear constraints and a fractional objective function. We use the approach of Charnes and Cooper (1962) to reformulate this optimization problem to obtain an equivalent linear program.

In order to apply the reformulation method, one has to ensure that the denominator term does not change its sign over the whole feasible space of the program to be reformulated. This requirement is fulfilled for Model (10): If we assume, that there is a feasible solution $\bar{\pi}$ of Model (10), for which the denominator is not positive, we can derive

$$\begin{aligned} 0 &\geq \bar{\pi}^T H(\tilde{x} - \bar{x}) - f^T(\tilde{x} - \bar{x}) \\ &= \bar{\pi}^T(b - H\bar{x}) + f^T\bar{x} - (\bar{\pi}^T(b - H\tilde{x}) + f^T\tilde{x}) \\ &> \beta - \beta = 0. \end{aligned}$$

The last inequality can is implied by Constraint (10d) and the conditions of the Lemma.

Hence, we can reformulate Model (10) by introducing a new variable $\alpha > 0$ and the constraint

$$\alpha = \frac{1}{\pi^T H(\tilde{x} - \bar{x}) - f^T(\tilde{x} - \bar{x})} \Leftrightarrow \alpha(\pi^T H(\tilde{x} - \bar{x}) - f^T(\tilde{x} - \bar{x})) = 1.$$

The objective function gets

$$\alpha(\pi^T(b - H\bar{x}) + f^T\bar{x} - \beta).$$

Further, we can multiply Constraints (10b) to (10d) with α. Substituting $\alpha\pi = \rho$, relaxing $\alpha > 0$ to $\alpha \geq 0$ and dropping the reformulation of Constraint (10d) yields

$$\rho^T(b - H\bar{x}) \geq \alpha(\beta - f^T\bar{x}).$$

This constraint bounds the objective function from below and is hence not necessary. Taking all together, we get Model (11).

For ii), it remains to show, that an optimal solution $(\bar{\rho}, \bar{\alpha})$ of Model (11) has the property $\bar{\alpha} > 0$. Assume, $\bar{\alpha} = 0$. Then, we know that

$$\bar{\rho}^T(b - H\bar{x}) \leq \begin{cases} \max \pi^T(b - H\bar{x}) \\ \text{s.t. } \pi^T A \leq 0 \\ \pi \geq 0. \end{cases}$$

The optimal value is 0, since we assumed $S(\bar{x}) < \infty$. Nevertheless, there is always a solution $\tilde{\pi}$ of Model (3) implying a cut $\tilde{\omega}$ that has the property that $\tilde{\omega}(\bar{x}) = S(\bar{x})$. If we define

$$\tilde{\alpha} := \frac{1}{\tilde{\pi}^T H(\tilde{x} - \bar{x}) - f^T(\tilde{x} - \bar{x})}, \ \tilde{\rho} := \tilde{\alpha}\tilde{\pi},$$

This is a feasible solution of Model (11) with a positive objective value - a contradiction. □

One can directly obtain results for the line shift procedure including only feasibility cuts, which are presented next.

Observation 2.12. *Given Problem* (1) *and the corresponding master problem* (2) *with feasible set and points* $\tilde{x} \in X$, $\bar{x} \in X$, *with* $S(\tilde{x}) < \infty$ *and* $S(\bar{x}) = \infty$. *A solution* $(\bar{\pi}, \bar{\mu})$ *of the following optimization problem is a line shift induced by* $(\tilde{x}, \bar{x}, \beta)$, *for all* $\beta \in \mathbb{R}$, *with depth* $\bar{\mu}$, *and no feasibility cut induced by this triple has a higher depth.*

$$\max \quad \mu \tag{12a}$$
$$s.t.\ \pi^T(b - H(\bar{x} + \mu(\tilde{x} - \bar{x}))) = 0 \tag{12b}$$
$$\pi^T A \leq 0 \tag{12c}$$
$$\pi^T(b - H\bar{x}) = 1 \tag{12d}$$
$$\pi \geq 0,\ \ \mu \geq 0. \tag{12e}$$

Especially, the optimization problem is bounded and has at least one feasible solution.

The following Lemma states that Model (12) can be reformulated similar to Model (10).

Lemma 2.13. *Given Problem* (1) *and the corresponding master problem* (2) *with feasible set and points* $\tilde{x} \in X$, $\bar{x} \in X$, *with* $S(\tilde{x}) < \infty$ *and* $S(\bar{x}) = \infty$. *Then, each optimal solution* $\bar{\rho}$ *of*

$$\max \quad \rho^T(b - H\bar{x}) \tag{13a}$$
$$\rho^T H(\tilde{x} - \bar{x}) = 1 \tag{13b}$$
$$\rho^T A \leq 0 \tag{13c}$$
$$\rho \geq 0. \tag{13d}$$

has the property, that $\bar{\pi} := \frac{\bar{\rho}}{\bar{\rho}^T(b - H\bar{x})}$, $\bar{\mu} := \bar{\rho}^T(b - H\bar{x})$ *is an optimal solution of Model* (12).

Proof. It is easy to check, that Constraint (12b) is equivalent to

$$\mu = \frac{\pi^T(b - H\bar{x})}{\pi^T H(\tilde{x} - \bar{x})}$$

Hence, μ can be replaced in the objective function of Model (12). The result is an optimization problem with linear constraints and a fractional objective function. We use the approach of Charnes and Cooper (1962) to reformulate this optimization problem to obtain an equivalent linear program.

That the reformulation method is applicable, it has to be guaranteed that the denominator term does not change its sign over the whole feasible space of the program to be reformulated. This requirement is fulfilled for Model (12): If we assume, that there is a feasible solution $\bar{\pi}$ of Model (12), for which the denominator is not positive, we can derive

$$\begin{aligned} 0 &\geq \bar{\pi}^T H(\tilde{x} - \bar{x}) \\ &= \bar{\pi}^T(b - H\bar{x}) - \bar{\pi}^T(b - H\tilde{x}) \\ &\geq 1 - 0 = 1. \end{aligned}$$

The last inequality can be derived from Constraint (12d) and the conditions of the Lemma.

Hence, we can reformulate Model (12) by introducing a new variable $\alpha > 0$ and the constraint

$$\alpha = \frac{1}{\pi^T H(\tilde{x} - \bar{x})} \Leftrightarrow \alpha\pi^T H(\tilde{x} - \bar{x}) = 1.$$

The objective function gets

$$\alpha\pi^T(b - H\bar{x}) = \alpha.$$

Further, we can multiply Constraints (12b) to (12d) with α. Substituting $\alpha\pi = \rho$, and substituting α by $\rho^T(b - H\bar{x})$ according to Constraint (12d) leads to Model (13). An optimal solution of Model (11) has a positive objective value. $\square$

We note that Model (13) is exactly Model (11) with α restricted to be 0. Hence, next lemma connects the results regarding feasibility cuts and optimality cuts.

Lemma 2.14. *Given Problem* (1) *and the corresponding master problem* (2) *with feasible set* X*, a budget* $\beta \in \mathbb{R}$ *and points* $\tilde{x} \in X$*,* $\bar{x} \in X$*, with* $f^T\bar{x} + S(\bar{x}) \geq \beta$*,* $f^T\tilde{x} + S(\tilde{x}) \leq \beta$*. If no optimality cut* $\bar{\omega}$ *with* $\bar{\omega}(\tilde{x}) + f^T\tilde{x} = \beta$ *and* $\bar{\omega}(\bar{x}) + f^T\bar{x} = \beta$ *exists, and if for an optimal solution* $(\bar{\rho}, \bar{\alpha})$ *of Model* (11) $\bar{\alpha} = 0$ *holds, then the optimal value of Model* (10) *is not greater than the optimal value of Model* (12)*.*

Proof. Assume the optimal value of Model (10) is $\bar{\mu}$, attained at $\bar{\pi}$, and the optimal value of Model (12) is $\tilde{\mu}$ with $\bar{\mu} > \tilde{\mu}$. Then, the constraint

$$\pi^T(b - H\bar{x}) \leq \bar{\pi}^T(b - H\bar{x})$$

can be added to Model (10) without changing the optimal value. This corresponds to an alternative dual sub problem which definitely has no unbounded rays. Lemma 2.11 ascertains that in this case, for an optimal solution $(\tilde{\rho}, \tilde{\alpha})$ of the modified sub problem, Model (11), it holds that $\tilde{\alpha} > 0$, with an optimal value of $\bar{\mu}$. Since $(\bar{\rho}, \bar{\alpha})$ is an optimal solution of the original Model (11), its value is hence at least $\bar{\mu}$. Since $\bar{\rho}$ is feasible for Model (13) and the optimal values of Model (12) and Model (13) coincide, we obtain

$$\tilde{\mu} \geq \bar{\mu} > \tilde{\mu},$$

a contradiction. $\square$

Taking Lemmas 2.11, 2.13 and 2.14 together, we get the following result.

Theorem 2.15. *Given Problem* (1) *and the corresponding master problem* (2) *with feasible set* X*, a budget* $\beta \in \mathbb{R}$ *and points* $\tilde{x} \in X$*,* $\bar{x} \in X$*, with* $f^T\bar{x} + S(\bar{x}) \geq \beta$*,* $f^T\tilde{x} + S(\tilde{x}) \leq \beta$*. If no optimality cut* $\bar{\omega}$ *exists for which* $\bar{\omega}(\tilde{x}) + f^T\tilde{x} = \beta$ *and* $\bar{\omega}(\bar{x}) + f^T\bar{x} = \beta$*, then each optimal solution* $(\bar{\rho}, \bar{\alpha})$ *of Model* (11) *defines a line shift with maximal depth. The depth equals the objective value of the model, and the cut that is added to Model* (2) *reads*

$$\bar{\rho}^T(b - Hx) \leq \bar{\alpha}\eta.$$

Proof. A line shift with maximal depth is either a feasibility cut, which is captured by Model (12), or an optimality cut, captured by Model (10), depending on which model has the higher optimal value. If we consider a solution $(\bar{\rho}, \bar{\alpha})$ of Model (11) with $\bar{\alpha} = 0$, Lemma 2.14 guarantees that the optimal value of Model (12) is not lower than the optimal value of Model (10). Furthermore, $\bar{\rho}$ is feasible for Model (13), and we can apply Lemma 2.13 to transform this solution into a feasibility cut, that reads

$$\bar{\rho}^T(b - Hx) \leq 0.$$

If $\bar{\alpha} > 0$, the solution is also a solution to Model (11) with α restricted to be positive. We further know that the optimal value of this model is not lower than the solution of Model (13), since it is a relaxation of the latter. The solution $(\bar{\rho}, \bar{\alpha})$ can be transformed by applying Lemma 2.11 to a solution of Model (10) with the same objective function value, and this can be transformed into a valid optimality cut that reads

$$\bar{\rho}^T(b - Hx) \leq \bar{\alpha}\eta.$$

Taking all together, we get the desired result. $\square$

The following lemma states, that Model (11) gets infeasible, if the requirements of Lemma 2.11 are not met, i.e., if $S(\tilde{x}) + f^T\tilde{x} > S(\bar{x}) + f^T\bar{x}$ for the current iterate $\bar{x}$ and incumbent $\tilde{x}$.

Lemma 2.16. *Given Problem* (1) *and the corresponding master problem* (2) *with feasible set X, a budget $\beta \in \mathbb{R}$ and points $\tilde{x} \in X$, $\bar{x} \in X$, with $f^T\bar{x} + S(\bar{x}) < \beta$, $f^T\tilde{x} + S(\tilde{x}) \leq \beta$. Then, Model* (11) *is either infeasible or has a negative optimal value.*

Proof. Assume, Model (11) has an optimal value that is positive, or is unbounded. Then it has a feasible solution $(\bar{\rho}, \bar{\alpha})$ with a positive value. If $\bar{\alpha} = 0$, we can derive that $\bar{\rho}^T A \leq 0$ and $\bar{\rho}^T(b - H\bar{x}) > 0$. This translates into a valid feasibility cut that cuts off $\bar{x}$, and this is a contradiction. If $\bar{\alpha} > 0$, we can derive that $\frac{1}{\bar{\alpha}}\bar{\rho}^T A \leq c^T$, and $\frac{1}{\bar{\alpha}}\bar{\rho}^T(b - H\bar{x}) + f^T\bar{x} > \beta$. This translates into a valid optimality cut, that guarantees that $f^T\bar{x} + S(\bar{x}) > \beta$, and this is a contradiction as well. □

The construction of OLS relies on the knowledge of good feasible solutions of the original optimization problems. We observed that it is beneficial, if the feasible solution the approach relies on is varied in every iteration. This is the motivation of the following corollary, stating how a new point in the convex hull of the master problem's feasible space can be constructed after an OLS cut is generated.

Corollary 2.17. *Given Problem* (1) *and the corresponding master problem* (2) *with feasible set X, a budget $\beta \in \mathbb{R}$ and points $\tilde{x} \in X$, $\bar{x} \in X$, with $f^T\bar{x} + S(\bar{x}) \geq \beta$, $f^T\tilde{x} + S(\tilde{x}) \leq \beta$. If no optimality cut $\bar{\omega}$ exists for which $\bar{\omega}(\tilde{x}) + f^T\tilde{x} = \beta$ and $\bar{\omega}(\bar{x}) + f^T\bar{x} = \beta$, $\hat{x} := \bar{x} + \bar{\mu}(\tilde{x} - \bar{x})$, where $\bar{\mu} < 1$ denotes the optimal value of Model* (11) *corresponding to this setting, then, it holds that*

$$\begin{aligned} &f^T\hat{x} + S(\hat{x}) \leq \beta, \\ &f^T x + S(x) < \beta \qquad \textit{for all } x \in \{\hat{x} + \lambda(\tilde{x} - \hat{x}), \lambda \in (0,1)\}. \end{aligned}$$

Proof. First, we show $f^T\hat{x} + S(\hat{x}) \leq \beta$. To do this, assume first that $S(\hat{x}) = \infty$. Then, there exists a valid feasibility cut $\tilde{\omega}$, such that $\tilde{\omega}(\hat{x}) > 0$. Since the cut is valid, it holds that $\tilde{\omega}(\tilde{x}) \leq 0$. Hence, there is $\hat{\mu} \in (0,1)$ for which $\tilde{\omega}(\bar{x} + \hat{\mu}(\tilde{x} - \bar{x})) = 0$ and $\tilde{\omega}(\bar{x} + \mu(\tilde{x} - \bar{x})) > 0$ for all $\mu \in (0, \hat{\mu})$. Hence, $\tilde{\omega}$ implies a solution of Model (11) with value $\hat{\mu} > \bar{\mu}$, and this is a contradiction to our assumption.

Assume, that $\infty > f^T\hat{x} + S(\hat{x}) > \beta$. Then, there exists a valid optimality cut $\tilde{\omega}$ such that $f^T\hat{x} + \tilde{\omega}(\hat{x}) > \beta$. Since the cut is valid, it holds that $f^T\tilde{x} + \tilde{\omega}(\tilde{x}) \leq \beta$. Hence, there is $\hat{\mu} \in (0,1)$, such that $f^T(\bar{x} + \hat{\mu}(\tilde{x} - \bar{x})) + \tilde{\omega}(\bar{x} + \hat{\mu}(\tilde{x} - \bar{x})) = \beta$ and $f^T(\bar{x} + \mu(\tilde{x} - \bar{x})) + \tilde{\omega}(\bar{x} + \mu(\tilde{x} - \bar{x})) > \beta$ for all $\mu \in (0, \hat{\mu})$. Hence, $\tilde{\omega}$ implies a solution of Model (11) with value $\hat{\mu} > \bar{\mu}$, and this is a contradiction.

Second, we show $f^T x + S(x) < \beta$ for all $x \in \{\hat{x} + \lambda(\tilde{x} - \hat{x}), \lambda \in (0,1)\}$. Assume, that there exists $\breve{x} \in \{\hat{x} + \lambda(\tilde{x} - \hat{x}), \lambda \in (0,1)\}$, i.e., $\breve{x} = \hat{x} + \breve{\lambda}(\tilde{x} - \hat{x})$ for a $\breve{\lambda} \in (0,1)$, with $f^T\breve{x} + S(\breve{x}) \geq \beta$. Then there is a valid optimality cut $\tilde{\omega}$ with $f^T\breve{x} + \tilde{\omega}(\breve{x}) \geq \beta$. Since this cut is valid, it holds that $f^T\hat{x} + \tilde{\omega}(\hat{x}) \leq \beta$ and $f^T\tilde{x} + \tilde{\omega}(\tilde{x}) \leq \beta$. Since one condition of the lemma was, that no optimality cut $\bar{\omega}$ exists for which $\bar{\omega}(\tilde{x}) + f^T\tilde{x} = \beta$ and $\bar{\omega}(\bar{x}) + f^T\bar{x} = \beta$, we can derive that

$$\beta > \breve{\lambda}(f^T\hat{x} + \tilde{\omega}(\hat{x})) + (1 - \breve{\lambda})(f^T\tilde{x} + \tilde{\omega}(\tilde{x})) = f^T\breve{x} + S(\breve{x}) \geq \beta,$$

and this is a contradiction. □

Remark. On the conditions of Theorem 2.15. We want to note that none of the conditions of Theorem 2.15 has to be checked before Model (11) is set up and solved. Since $\tilde{x}$ is supposed to correspond to be the best feasible solution of Model (1) found so far, and β is its optimal value, $f^T\tilde{x} + S(\tilde{x}) \leq \beta$ is trivial. Lemma 2.16 ascertains, that whenever $f^T\bar{x} + S(\bar{x}) < \beta$, i.e., $\bar{x}$ improves the currently best known objective value of Model (1), this reflects in the solution process of Model (11) in any case. Hence, this condition has not to be checked upfront. In contrast, if a valid optimality cut $\bar{\omega}$ exists, for which $\bar{\omega}(\tilde{x}) + f^T\tilde{x} = \beta$ and $\bar{\omega}(\bar{x}) + f^T\bar{x} = \beta$, it is not implied by a feasible solution of Model (11), since such a cut would let the left-hand side of Constraint (11b) be equal to 0. Nevertheless, such a cut is preferable, since it is a line shift maximum depth. However, solving Model (11) results in any case either in a negative optimal value or infeasibility. We proceed with solving the original dual sub problem, or a non-negative optimal value between 0 and 1, that implies a cut that cuts off the current iterate, and we simply insert it. Alternatively, the condition can easily be checked on the fly in an algorithmic implementation of the method, and under the assumption that we have a point $\tilde{x}$ at hand with the property that $f^T\tilde{x} + S(\tilde{x}) < \beta$, it holds in any case. Corollary 2.17 states how such points can be obtained.

Remark: Connection of facet cuts and optimal line shifts. The OLS strategy can also be considered as a facet generating strategy, with the difference, that instead of the point $(\bar{x}, \bar{\eta})$ the point $(\bar{x}, \beta - f^T\bar{x})$ is separated from the epigraph.

Remark: Sub problems with block-diagonal structure. BD works exceptionally well, if the sub problem (3) decomposes into several independent optimization problems. This is the case, if the sub problem has a so-called block-diagonal structure. The independent optimization problems can then be solved separately, which leads to a considerably reduced computational effort.

Block-diagonal structure is, strictly speaking, not immediately compatible with the procedure introduced within this chapter, since the surrogate problems do not decompose into blocks anymore (Constraint (11b) couples the blocks).

Nevertheless, we want to superficially propose a strategy how an existing block-diagonal structure can be exploited within the provided approach. The idea is based on the last remark, i.e., the fact that we separate the point $(\bar{x}, \beta - f^T\bar{x})$ from the epigraph of S in the sense of Stursberg (2019), whenever this is possible, i.e., if $\beta - f^T\bar{x} < S(\bar{x})$. This condition can easily be checked by optimizing the duals of the single blocks simultaneously, until the cumulated solution value reaches $\beta - f^T\bar{x}$. A value $\bar{\eta}$ is obtained for each block of the sub problem, such that $(\bar{x}, \bar{\eta})$ is not in the epigraph of the value function of the corresponding block. Hence, this point can be separated from this epigraph, optimizing only over a single block of the dual sub problem. It is not guaranteed that this procedure delivers a line shift with maximal depth for the overall optimization problem. Nevertheless, some properties of the original approach are pertained: The current iterate $\bar{x}$ is excluded from further consideration, and in certain cases a set of points can be determined, for which the value of the overall problem is lower than β.

The following chapter specifies the algorithmic framework we propose to solve optimization problems with the approach provided in this chapter.

3. The algorithmic framework.

In this section, we discuss different solution algorithms. First, we explain how the OLS procedure is implemented.

The implementation of the methods Magnanti-Wong, MIS and facet is quite straight-forward: The corresponding sub problems presented in Section 2.2 are created, solved, and the corresponding cut is added to the master problem. In contrast, Theorem 2.15 can only be applied if the current master solution has a value that is above the value of the currently best known solution. Furthermore, the theorem has as the condition, that no valid optimality cut $\bar{\omega}$ exists that has the property that $f^T\bar{x}+\bar{\omega}(\bar{x}) = f^T\tilde{x}+\bar{\omega}(\tilde{x}) = \beta$. Hence we elaborate more on frameworks how this condition can be handled.

We first note that it can easily be checked if there is a valid optimality cut $\bar{\omega}$ that has the property that $f^T\bar{x} + \bar{\omega}(\bar{x}) = f^T\tilde{x} + \bar{\omega}(\tilde{x}) = \beta$. We only need to solve Model (3) with the additional constraint

$$\pi^T(b - H\tilde{x}) = \beta - f^T\tilde{x}. \tag{14}$$

Since it is clear that this model is infeasible if $f^T\tilde{x} + S(\tilde{x}) < \beta$, this step can be omitted if we know this to hold. As pointed out in the first remark in the end of the last section, this step is optional.

There are different ways to continue. One could solve Model (3), depending on the current master solution $\bar{x}$, but only until a feasible solution, that has a value that exceeds $\beta - f^T\bar{x}$, or unboundedness is detected. In both cases, this guarantees the conditions of Theorem 2.15. Furthermore, the solution/ray obtained in the incomplete optimization process of Model (3) can be completed to a feasible solution of Model (11).

In the first case, we receive a vector $\bar{\pi}$, with the property that

$$\begin{aligned} \bar{\pi}^T(b - H\bar{x}) &\geq \beta - f^T\bar{x} \\ \bar{\pi}^T A &\leq c^T \end{aligned}$$

Using the same argument as in the proof of Lemma 2.11, we state that

$$\psi := \bar{\pi}^T H(\tilde{x} - \bar{x}) - f^T(\tilde{x} - \bar{x}) \geq 0,$$

with equality, if and only if

$$\omega_{\bar{\pi}}(\bar{x}) + f^T\bar{x} = \omega_{\bar{\pi}}(\tilde{x}) + f^T\tilde{x}.$$

If this holds, $\bar{\pi}$ induces an OLS. If it is not the case, we set $\bar{\rho} := \frac{1}{\psi}\bar{\pi}$, and $\bar{\alpha} := \frac{1}{\psi}$ to obtain a feasible solution of Model (11). We use this as a start solution and continue optimizing Model (11).

We note that it is also be possible to solve Model (3) first to optimality to check the conditions of Theorem 2.15. Nevertheless, we made the experience that the conditions to apply Theorem 2.15 seem to be met in the majority of iterations. This implies, that firstly solving Model (3) results in having to solve two LPs in the majority of iterations. In contrast, we can apply Lemma (2.16), that ascertains that the Model (11) gets infeasible or has a negative optimal value if and only if the conditions of Theorem 2.15 are not met. This is the second framework we propose. It solves Model (11) first. Afterwards, it is observed if it is infeasible or has a negative optimal value. If this is the case, we solve Model (3) to obtain the true value of $S(\bar{x})$. With this strategy, we have to solve one LP in the majority of iterations, and we only have to solve two LPs in very few iterations.

We want to benchmark our cut selection strategy against the existing ones. This includes the Magnanti-Wong method, MIS cut selection, and facet generating cuts. Since we test the algorithms on mixed-integer linear programs, with the integer variables in the master problem and the continuous variables in the sub problem, we implemented a Branch-and-Benders-Cut procedure, that starts a single Branch-and-Cut framework, and generates valid cuts whenever a new integer solution of the master problem is generated. It maintains values for $x^* \in X$, $\tilde{x} \in \text{conv}(X)$, and $\beta \in \mathbb{R}$. Algorithm 1 shows which steps are conducted whenever the Branch-and-Cut framework detects a new integral solution that has a value that is lower than the value of the best solution found so far. The Branch-and-Cut framework inserts the returned cut ω and updates the values of $\tilde{x}, x^*$ and β afterwards. It terminates, as soon as "optimal" is returned, with a solution x^* that is part of a solution of Model (1), that is at most $tol \in \mathbb{R}$ worse than an optimal one.

Algorithm 1 OLS Benders cut generation

Input $\bar{x}, x^* \in X$, $\bar{\eta} \in \mathbb{R}$, $\tilde{x} \in \operatorname{conv}(X)$, with $\infty \neq \beta := f^T x^* + S(x^*) \geq f^T \tilde{x} + S(\tilde{x})$, $(\bar{x}, \bar{\eta})$ integral solution of (2) with $f^T \bar{x} + \bar{\eta} < \beta$, and a lower bound l on (2).
Output Message, that (1) is solved **or** a valid cut and updated $\tilde{x}, x^*, \beta$.

```
if l ≤ β − tol then
    Solve (11). Let μ̄ denote its optimal value.
    if Model (11) is infeasible or μ̄ < 0 then
        Solve Model (3).
        ω ← Benders cut implied by (3).
        β ← optimal value of (3).
        x̃, x* ← x̄.
    else
        ω ← Benders cut implied by (11).
        x̃ ← x̃ ∈ {x̄ + μ(x̃ − x̄), μ ∈ [μ̄, 1]}.
    end if
    return x̃, x*, β, ω.
else
    return optimal.
end if
```

For the cut selection methods that require a feasible solution of the optimization problem to get started, i.e., facet and OLS, a feasible solution calculated using a heuristic is provided. The time required to execute this heuristic does not contribute to the solution time of the methods. In the following section, the results of our computational study are presented.

4. Computational Results

In this section, we present our computational results. This includes a description of our computational setup, and the presentation of results for different classes of instances: Instances taken from the MIPLib, multicommodity-flow network design instances, and instances of randomly generated MIPs. All instances and our code are publicly available, see Glomb et al. (2023a).

4.1. **Computational Setup.** In this section, the results of our computational study are presented. All programs have been written using the programming language Python, version 3.10. Optimization problems have been solved in single-thread mode using the Gurobi optimizer, version 10.0.2, see Gurobi Optimization (2020). The programs have been executed on nodes of a high performance computing cluster using a Xeon E3-1240 v6 CPU with four cores at 3.7GHz base frequency and with a total memory of 32 GB. We solved four instances on one node at a time. All calculations are terminated after one hour, leaving the processes 180 seconds for non-optimization tasks like setting up the problems or writing out the solution files. The duration of these tasks has not been included in the reported solution times. Whenever a relative optimality gap of 10^{-4} has been reached, the calculation has been terminated as well. Since we only consider minimization problems with a positive optimal value, the optimality gap is defined depending on a known upper bound $u > 0$ on optimal value and a known lower bound $u \geq l \geq 0$ on the optimal value as $\frac{u-l}{u} \in [0, 1]$. All instances have been solved directly with Gurobi without using a decomposition approach first (**van**). Furthermore, we tried BD using five different cut selection approaches: the Magnanti-Wong strategy (**mwb**), the MIS strategy (**mis**), the Facet strategy (**fcb**), the OLS strategy (**ols**), and a hybrid strategy (**hyb**) combining MIS and OLS. The hybrid strategy uses

Instance set	Facet	Magnanti-Wong	MIS	OLS	MIP-Solver	Hybrid	Benders	Total
MIPLib	19	18	18	18	39	18	19	43
MCF-NWD	24	16	26	26	37	26	26	50
Fischetti et al. (2010)	19	18	18	19	12	19	19	20
Random MIP	45	46	45	48	48	45	48	48

Table 1. Instances per instance set that could be solved to optimality by different solution approaches within the time limit. The right column denotes the size of the instance set.

MIS selection for at least 180 seconds, and at least until the optimality gap reaches 10%. Then it uses the OLS selection. We want to note that the hybrid strategy has the additional advantage that it can be started without knowing a feasible solution. For the cut selection strategies, that rely on a feasible solution to be known (Facet and OLS), we provided a feasible solution, that has an objective within 25% of the optimal value. All instances are mixed-integer linear programs, and all integer variables have been retained in the master problem, while all continuous variables have been put into the sub problem. All cut selection strategies could hence be embedded into a Branch-and-Cut framework, analogous to the framework described in Algorithm 1.

In order to establish a proper graphical illustration of the algorithms' performance, performance plots for the algorithms' solution time, number of required cuts and optimality gap after one hour have been generated. Cut performance plots and time performance plots have been generated including only instances, that could be solved to optimality within the time limit by at least one algorithm that is based on BD. This prevents that the evaluations are strongly biased by assigning the algorithm that produces new cuts at the slowest rate overly high cut performance values. Table 1 shows how many instances per instance set could be solved to optimality by the different solution approaches. The second column from right denotes how many instances could be solved to optimality by BD for at least one of the proposed cut selection strategies. These instances have been used to generate the performance plots for the runtime of the algorithms. The runtime has been set to the time limit, whenever an instance has not been solved to optimality. Hence, the performance plots that are shown might overestimate the true performance plots by $1 - \frac{\text{\# Instances solved by algorithm}}{\text{\# Instances solved by Benders}}$. The same applies to the cut efficiency plots. Whenever a significant proportion of instances could not be solved to optimality, we additionally presented a performance plot for the optimality gap of these instances. For the other instance sets, the gaps for instances that could not be solved using BD are reported in tables whenever suitable.

In the following, we will present our results on MIPLIB instances, network design problems and randomly generated MIPs. The results are depicted as so-called performance plots. Performance plot graphs are defined by the monotone functions

$$\phi_a : [1, \infty) \to [0, 1], \quad a \in A,$$
$$\rho \mapsto \frac{|\{i \in I : p(a, i) \leq \rho \min_{\tilde{a} \in A} p(\tilde{a}, i)\}|}{|I|},$$

where I is a set of instances, A is a set of algorithms and p is a performance measure, depending on algorithm and instance, like, as in our case, "time the algorithm needs to solve the instance", "cuts the algorithm needs to solve the instance" or "optimality gap after termination". The graphs of algorithms with high performance are in the top-left region of the grid.

4.2. **Results for MIPLib Instances.** We benchmarked our cut selection strategy against MIS, Facet and Magnanti-Wong cuts on a selection of decomposable instances from the MIPLib (2017) Collection set Gleixner et al. (2021). We identified 55 instances with the following properties:

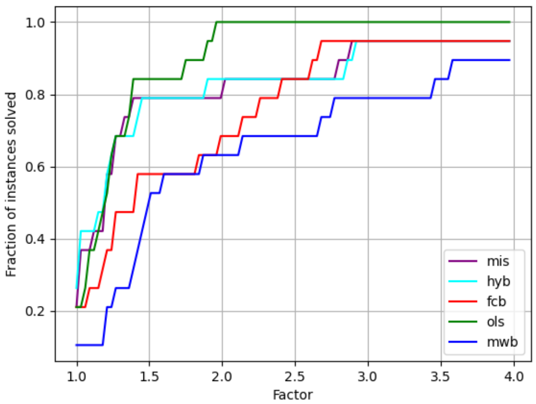

FIGURE 2. Fraction of MIPLIB instances that could be solved within a multiple of fastest running time.

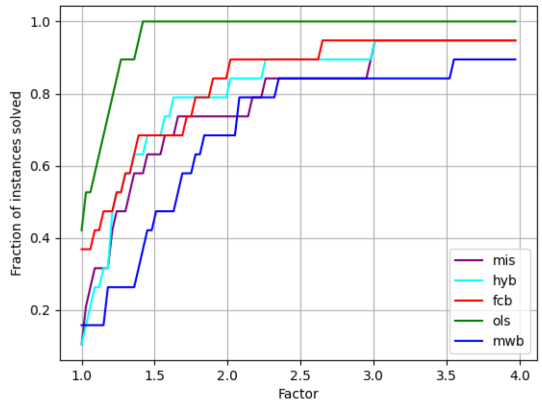

FIGURE 3. Fraction of MIPLIB instances that could be solved using a multiple of the cuts needed by the algorithm with the fewest cuts.

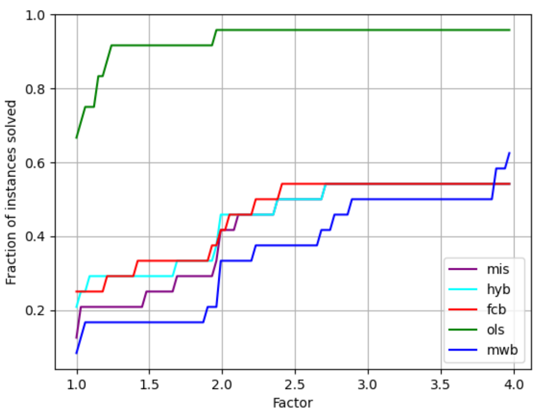

FIGURE 4. Fraction of MIPLIB instances achieving a multiple of the gap the algorithm with the lowest gap achieves after one hour of calculation.

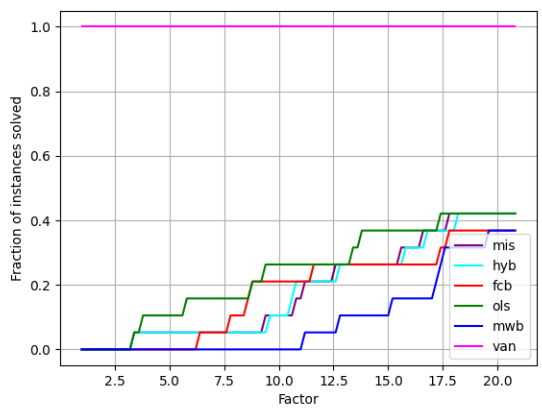

FIGURE 5. Fraction of MIPLIB instances that could be solved within a multiple of fastest running time, including the state-of-the-art MIP solver.

they should contain at least 1 and at most 1000 integer/binary variables, at least 1 and at most 100000 continuous variables, at most 10000 constraints. They should contain more continuous variables than integer variables, and have a positive objective function value. We only selected instances labeled as "easy" and not labeled as "numerics".

Of these 55 instances, we sorted out the 12 instances binkar10_1, dano3_5, exp-1-500-5-5, fastxgemm-n2r6s0t2, neos-3072252-nete, neos-3627168-kasai, neos-3665875-lesum, neos-480878, neos22, newdano, rentacar and uct-subprob due to numerical problems. The results for the remaining 43 instances are shown in Figures 2 to 5.

19 of 43 MIPLIB instances could be solved to optimality using BD. Figure 2 shows that, measured in time needed to solve the instances, the hybrid selection approach, the MIS selection approach and OLS selection are competitive.

Comparing the number of cuts needed to solve instances to optimality, as shown in Figure 3, OLS clearly dominates the other cut selection strategies, needing the fewest cuts for over 40% of all instances. For no instance it needs more than 50% more cuts to solve it to optimality, compared to the best selection strategy for this instance. We want to note that the hybrid strategy that has competitive solution times, is clearly the second best strategy in terms of cut efficiency.

Figure 4 shows that for the remaining 24 instances that could not be solved using BD, OLS achieves the lowest optimality gap for over 60% of all instances.

We are not very surprised that Figure 5 demonstrates that BD is, regardless of the cut selection strategy, not competitive against one of the best state-of-the-art MIP solvers on MIPLIB instances.

We can conclude, that the hybrid strategy, the MIS strategy and the OLS strategy are the best ones on the MIPLIB test set.

4.3. **Results for Network Design Problems.** We tested the algorithms 50 different instances that we created ourselves, and 20 of the original instances used in Fischetti et al. (2010) (those with positive flow costs, i.e., the "optimality" instances). The instances we created ourselves are network design problems, with all combinations of: 20, 40, 60, 80 or 100 commodities, graphs that have either 5×4 or 6×4 nodes, and edges that are created either as grid, as Erdös-Renyi-Graph, as random 5-regular graph, as graph that is initialized as empty and adds random edges until each node has a degree between 2 and 6, and as graph that is derived of a random 4-regular graph, removing each edge with a probability of 0.06 and adding 13 random edges afterwards. Setup costs for an arc have been set to 5, while costs for a unit of flow along an arc have been set to 1. Each commodity has one origin node and one destination node and a demand of 1. The algorithmic performance is shown in Figures 6 - 9.

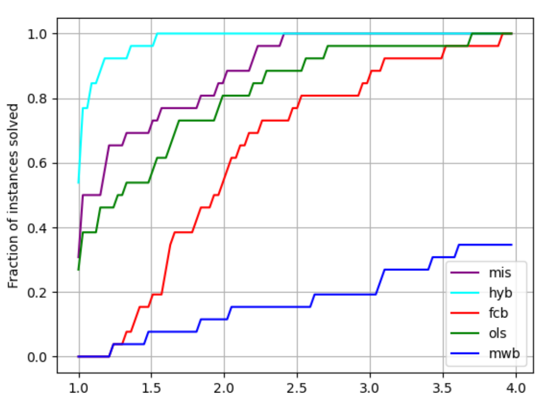

FIGURE 6. Fraction of network design instances that could be solved within a multiple of fastest running time.

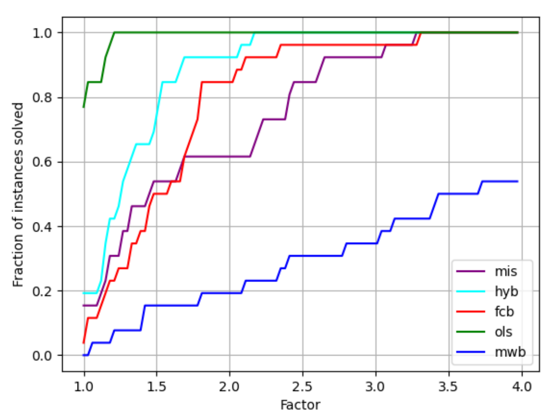

FIGURE 7. Fraction of network design instances that could be solved using a multiple of the cuts needed by the algorithm with the fewest cuts.

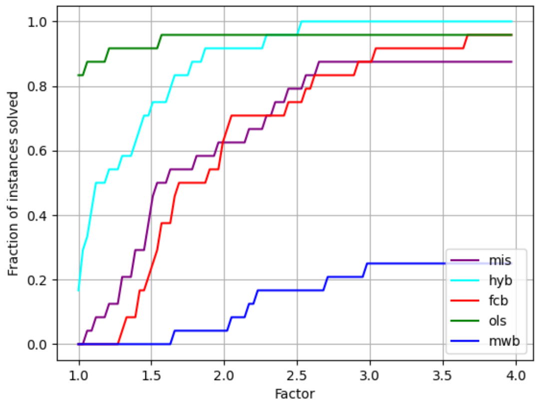

FIGURE 8. Fraction of network design instances achieving a multiple of the gap the algorithm with the lowest gap achieves after one hour of calculation.

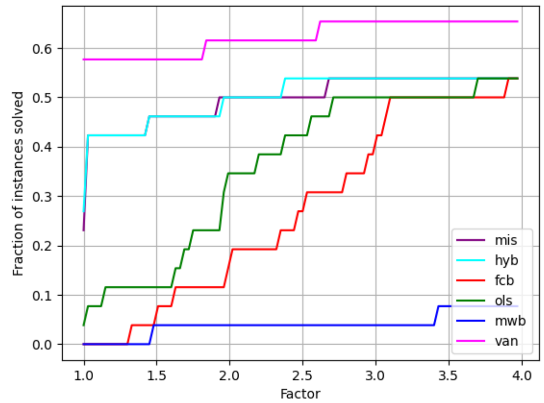

FIGURE 9. Fraction of network design instances that could be solved within a multiple of fastest running time, including the state-of-the-art MIP solver.

Figure 6 shows that in terms of solution time, the MIS strategy and the hybrid strategy are the competitive ones, while the hybrid strategy performs slightly better. These strategies are the optimal ones for approximately 70% of of the instances, while the hybrid selection strategy solves almost all instances within additional 60% of the best algorithm's running time.

Figure 7 shows that for the instances that could be solved to optimality by BD, the OLS strategy needs by far the fewest cuts.

Figure 8 shows that for the instances that could not be solved to optimality by one of the cut selection strategies, OLS selection clearly achieves the best optimality gaps after one hour.

Figure 9 shows that a state-of-the-art MIP solver is the best choice for almost 60% of all network design instances we created, while the hybrid approach/the MIS approach is the best for 40% of all instances tested. Pure OLS selection is the best approach for around 5% of the instances.

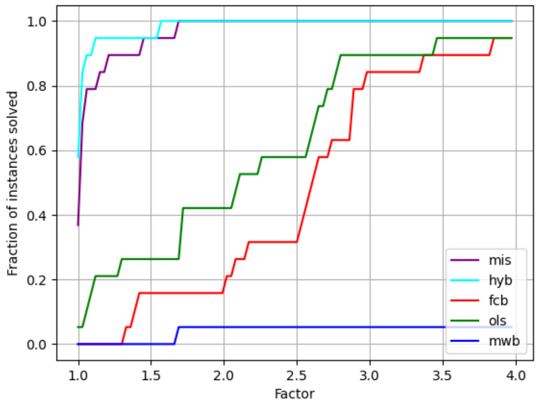

FIGURE 10. Fraction of Fischetti et al. (2010) instances that could be solved within a multiple of fastest running time.

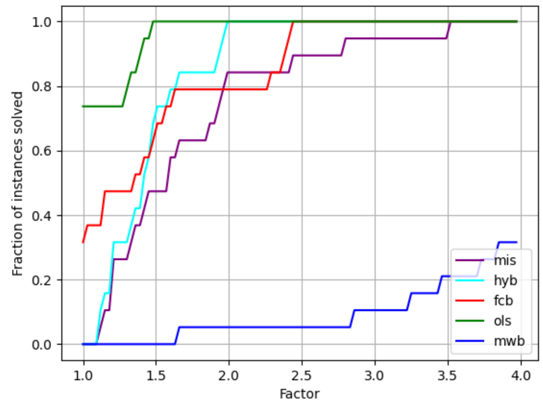

FIGURE 11. Fraction of Fischetti et al. (2010) instances that could be solved using a multiple of the cuts needed by the algorithm with the fewest cuts.

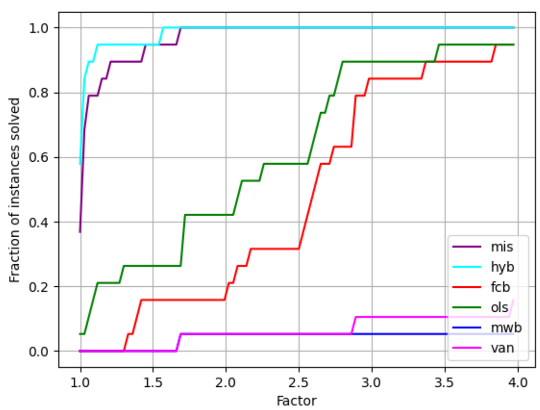

FIGURE 12. Fraction of Fischetti et al. (2010) instances that could be solved within a multiple of fastest running time, including the state-of-the-art MIP solver.

We want to note that the hybrid strategy and the Facet cuts are quite similar in their cut number performance, while MIS has a clearly worse performance. Hence, we can conclude that the hybrid strategy is competitive.

The results on the instances from Fischetti et al. (2010) are graphically summarized in Figures 10 - 12. As before, the hybrid strategy and MIS are competitive regarding time (Figure 10) and OLS and hybrid selection are competitive regarding cuts (Figure 11). For all instances from Fischetti et al. (2010) that have been tested, the classical, the state-of-the-art MIP solver is outperformed drastically as demonstrated in Figure 12.

4.4. **Results for randomly generated MIPs.** We present our computational results for randomly generated MIPs. The problems are generated randomly of all combinations of 50, 100 or 150 integer variables, 200 or 400 continuous variables, 50 or 100 inequality constraints containing only integer variables, 100 or 200 inequality constraints containing integer and continuous variables, and 200 or 400 inequality constraints containing only continuous variables. Additionally, each instance contains 5 equality constraints containing integer and continuous variables.

Integer variables have an objective function coefficient between 0.5 and 5. Continuous variables have an objective function coefficient between 0 and 10 times the ratio of the number of integer variables and the number of continuous variables. The objective function coefficients are drawn from these intervals following a uniform distribution. Each integer variable has a nonzero coefficient in constraints that contain integer variables with a probability that equals the ratio of 10 and the number of integer variables. Each continuous variable has a nonzero coefficient in constraints that contain continuous variables with a probability that equals the ratio of 20 and the number of integer variables.

The results can be taken from Figures 13 - 15.¸

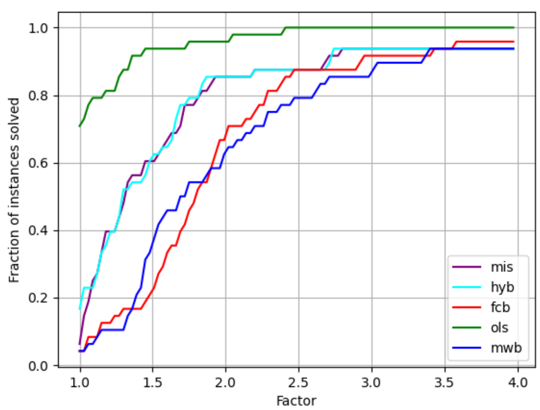

FIGURE 13. Fraction of Random MIP instances that could be solved within a multiple of fastest running time.

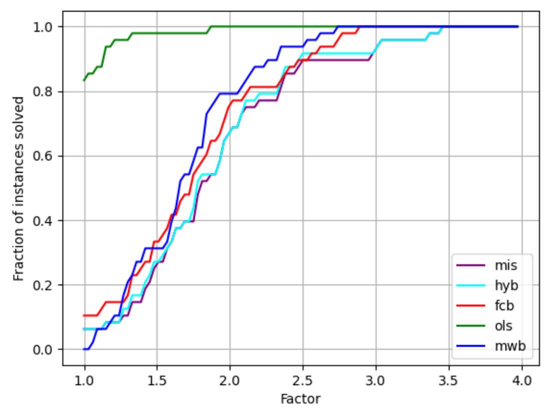

FIGURE 14. Fraction of Random MIP instances that could be solved using a multiple of the cuts needed by the algorithm with the fewest cuts.

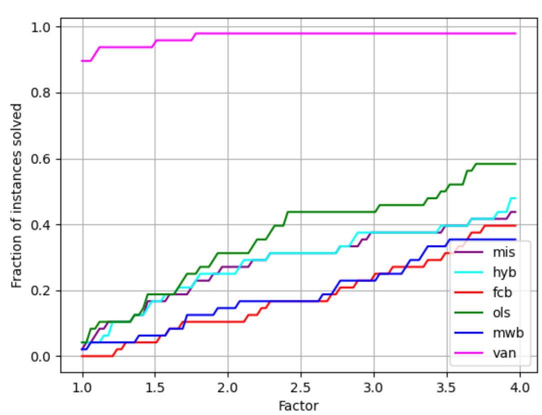

FIGURE 15. Fraction of Random MIP instances that could be solved within a multiple of fastest running time, including the state-of-the-art MIP solver.

	MIS	**Facet**	**OLS**
Adv.	• Practical performance is excellent • Easy to implement • No update of normalization constraint necessary • Works without knowledge of feasible solution	• Practical performance is high • Independent on constraint scaling and redundant constraints	• Practical performance is excellent • Immediate recognition of improvements • Independent on constraint scaling and redundant constraints • Always cuts off iterate point
Dis-adv.	• Depends on constraint scaling • Affected by redundant constraints • Difficult to interpret • Does not recognize improvements immediately • Needs sign-restricted dual variables • Does not necessarily cut off iterate	• Knowledge of feasible solution is necessary • Does not recognize improvements immediately • Update of normalization constraint in each iteration necessary • Does not necessarily cut off iterate	• Knowledge of feasible solution is necessary • In case of an improvement, two sub problems have to be solved • Update of normalization constraint in each iteration necessary

Table 2. Properties of MIS, OLS and Facet selection strategies.

Figure 13 implies that OLS, MIS and hybrid cut selection are competitive in terms of calculation time on this instance set. OLS is the fastest selection strategy for over 70% of all instances, and solves all instances in not more than 150% more time than the fastest algorithm.

Figure 14 shows that OLS cut selection is the only competitive when comparing the number of needed cuts. For over 80% of all instances, OLS needs the fewest cuts to solve the instance, and for no instance it needs more than twice as many cuts than the method needing the fewest cuts. Hence, OLS outperforms all other cut selection strategies.

Figure 15 demonstrates that for the random MIPs we generated, using a state-of-the-art MIP solver is superior to choosing BD as solution approach.

Since it has turned out that MIS, Facet and OLS cuts are the competitive ones, we conclude with a summary of what we consider to be the advantages and disadvantages of the three selection strategies in Table 2.

5. Summary and Outlook

In this article, first an innovative notion of Pareto-optimality for Benders cuts has been developed. This notion is based on so-called solution-candidate sets, that describe the set of points feasible for the master problem, that are potentially an optimal solution of the original optimization problem. We showed that cuts, that are non-dominated in our sense are also non-dominated in the sense of Magnanti and Wong (1981), but the opposite is not necessarily true.

Based on our notion of Pareto-optimality, we developed a novel cut selection strategy for BD, that is capable of calculating Pareto-optimal cuts if some mild conditions hold.

Further, we developed the algorithmic framework necessary to optimally exploit the potential of the cut selection strategy. The algorithm has been benchmarked against other known cut selection strategies (Magnanti-Wong, MIS, Facet) on various instance classes. For all instance classes (MIPLib, multi-commodity flow network design problems, randomly generated mixed-integer linear programming problems) the computational results show, that the developed method is competitive measured in CPU seconds needed to solve a problem to optimality, and the results showed that the developed method needs to generate a smaller number of cuts than the benchmark approaches to solve instances. The method is hence especially effective in situations with scarce memory or with a difficult to solve sub problem.

Possible future research directions are the transfer of the method to more general versions of the approach, like Generalized BD as published in Geoffrion (1972). The approach (as well as the MIS selection strategy as published in Fischetti et al. (2010) and Facet cuts as published in Stursberg (2019)) can be exploited to speed-up various algorithms based on BD that have been applied to solve real-world optimization problems, of which many are mentioned in Section 1, since in many of these articles, the approach is either applied without considering a cut selection strategy at all, or using the Magnanti-Wong method. Considering the high performance of MIS, Facet and OLS cuts for the optimization problems investigated in this article, this could lead to significant speedups for real-world applications as well.

Acknowledgements

This research was conducted within the framework of the research project OPs-TIMAL, financed by the German Federal Ministry of Economic Affairs and Energy (BMWi).

References

Abdolmohammadi, H. R. and Kazemi, A. (2013). A benders decomposition approach for a combined heat and power economic dispatch. *Energy conversion and management*, 71:21–31.

Adulyasak, Y., Cordeau, J.-F., and Jans, R. (2015). Benders decomposition for production routing under demand uncertainty. *Operations Research*, 63(4):851–867.

Ambrosius, M., Grimm, V., Kleinert, T., Liers, F., Schmidt, M., and Zöttl, G. (2020). Endogenous price zones and investment incentives in electricity markets: An application of multilevel optimization with graph partitioning. *Energy Economics*, 92:104879.

Azad, N., Saharidis, G. K., Davoudpour, H., Malekly, H., and Yektamaram, S. A. (2013). Strategies for protecting supply chain networks against facility and transportation disruptions: an improved benders decomposition approach. *Annals of Operations Research*, 210:125–163.

Baringo, L. and Conejo, A. J. (2011). Wind power investment: A benders decomposition approach. *IEEE Transactions on Power Systems*, 27(1):433–441.

Bärmann, A., Liers, F., Martin, A., Merkert, M., Thurner, C., and Weninger, D. (2015). Solving network design problems via iterative aggregation. *Mathematical Programming Computation*, 7:189–217.

Bayram, V. and Yaman, H. (2018). Shelter location and evacuation route assignment under uncertainty: A benders decomposition approach. *Transportation science*, 52(2):416–436.

Benders, J. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische mathematik*, 4(1):238–252.

Bonami, P., Salvagnin, D., and Tramontani, A. (2020). Implementing automatic benders decomposition in a modern mip solver. In *Integer Programming and Combinatorial Optimization: 21st International Conference, IPCO 2020, London, UK, June 8–10, 2020, Proceedings*, pages 78–90. Springer.

Botton, Q., Fortz, B., Gouveia, L., and Poss, M. (2013). Benders decomposition for the hop-constrained survivable network design problem. *INFORMS journal on computing*, 25(1):13–26.

Brandenberg, R. and Stursberg, P. (2021). Refined cut selection for benders decomposition: applied to network capacity expansion problems. *Mathematical Methods of Operations Research*, 94(3):383–412.

Charnes, A. and Cooper, W. W. (1962). Programming with linear fractional functionals. *Naval Research logistics quarterly*, 9(3-4):181–186.

Conforti, M. and Wolsey, L. A. (2019). "facet" separation with one linear program. *Mathematical Programming*, 178:361–380.

Contreras, I., Cordeau, J.-F., and Laporte, G. (2011). Benders decomposition for large-scale uncapacitated hub location. *Operations research*, 59(6):1477–1490.

Costa, A. M., Cordeau, J.-F., Gendron, B., and Laporte, G. (2012). Accelerating benders decomposition with heuristic master problem solutions. *Pesquisa Operacional*, 32:03–20.

Cote, G. and Laughton, M. A. (1984). Large-scale mixed integer programming: Benders-type heuristics. *European Journal of Operational Research*, 16(3):327–333.

Fischetti, M., Ljubić, I., and Sinnl, M. (2016). Benders decomposition without separability: A computational study for capacitated facility location problems. *European Journal of Operational Research*, 253(3):557–569.

Fischetti, M., Ljubić, I., and Sinnl, M. (2017). Redesigning benders decomposition for large-scale facility location. *Management Science*, 63(7):2146–2162.

Fischetti, M., Salvagnin, D., and Zanette, A. (2010). A note on the selection of benders' cuts. *Mathematical Programming*, 124(1):175–182.

Froyland, G., Maher, S. J., and Wu, C.-L. (2014). The recoverable robust tail assignment problem. *Transportation Science*, 48(3):351–372.

Gendron, B., Scutellà, M. G., Garroppo, R. G., Nencioni, G., and Tavanti, L. (2016). A branch-and-benders-cut method for nonlinear power design in green wireless local area networks. *European Journal of Operational Research*, 255(1):151–162.

Geoffrion, A. M. (1972). Generalized benders decomposition. *Journal of optimization theory and applications*, 10:237–260.

Geoffrion, A. M. and Graves, G. W. (1974). Multicommodity distribution system design by benders decomposition. *Management science*, 20(5):822–844.

Gleixner, A., Hendel, G., Gamrath, G., Achterberg, T., Bastubbe, M., Berthold, T., Christophel, P. M., Jarck, K., Koch, T., Linderoth, J., Lübbecke, M., Mittelmann, H. D., Ozyurt, D., Ralphs, T. K., Salvagnin, D., and Shinano, Y. (2021). MIPLIB 2017: Data-Driven Compilation of the 6th Mixed-Integer Programming Library. *Mathematical Programming Computation*.

Glomb, L., Liers, F., and Rösel, F. (2023a). Instances and code. `https://faubox.rrze.uni-erlangen.de/getlink/fi4cRCDxNtD7E5mwtS4m6R/`.

Glomb, L., Liers, F., and Rösel, F. (2023b). Optimizing integrated aircraft assignment and turnaround handling. *European Journal of Operational Research*, 310(3):1051–1071.

Glomb, L., Liers, F., and Rösel, F. (2023c). A stochastic optimization approach for optimal tail assignment with knowledge-based predictive maintenance. *CEAS Aeronautical Journal*, pages 1–14.

Grimm, V., Kleinert, T., Liers, F., Schmidt, M., and Zöttl, G. (2017). Optimal price zones of electricity markets: A mixed-integer multilevel model and global solution approaches. *Optimization Methods and Software*.

Gurobi Optimization, L. (2020). Gurobi optimizer reference manual.

Magnanti, T. L. and Wong, R. T. (1981). Accelerating benders decomposition: Algorithmic enhancement and model selection criteria. *Operations research*, 29(3):464–484.

Maheo, A., Kilby, P., and Van Hentenryck, P. (2019). Benders decomposition for the design of a hub and shuttle public transit system. *Transportation Science*, 53(1):77–88.

Mansouri, S., Ahmarinejad, A., Ansarian, M., Javadi, M., and Catalao, J. (2020). Stochastic planning and operation of energy hubs considering demand response programs using benders decomposition approach. *International Journal of Electrical Power & Energy Systems*, 120:106030.

McDaniel, D. and Devine, M. (1977). A modified benders' partitioning algorithm for mixed integer programming. *Management Science*, 24(3):312–319.

Nasri, A., Kazempour, S. J., Conejo, A. J., and Ghandhari, M. (2015). Network-constrained ac unit commitment under uncertainty: A benders' decomposition approach. *IEEE transactions on power systems*, 31(1):412–422.

Papadakos, N. (2008). Practical enhancements to the Magnanti–Wong method. *Operations Research Letters*, 36(4):444–449.

Poojari, C. A. and Beasley, J. E. (2009). Improving benders decomposition using a genetic algorithm. *European Journal of Operational Research*, 199(1):89–97.

Qian, L. P., Zhang, Y. J. A., Wu, Y., and Chen, J. (2013). Joint base station association and power control via benders' decomposition. *IEEE Transactions on Wireless Communications*, 12(4):1651–1665.

Rahmaniani, R., Crainic, T. G., Gendreau, M., and Rei, W. (2017). The benders decomposition algorithm: A literature review. *European Journal of Operational Research*, 259(3):801–817.

Rahmaniani, R., Crainic, T. G., Gendreau, M., and Rei, W. (2018). Accelerating the benders decomposition method: Application to stochastic network design problems. *SIAM Journal on Optimization*, 28(1):875–903.

Rei, W., Cordeau, J.-F., Gendreau, M., and Soriano, P. (2009). Accelerating benders decomposition by local branching. *INFORMS Journal on Computing*, 21(2):333–345.

Saharidis, G. K., Boile, M., and Theofanis, S. (2011). Initialization of the benders master problem using valid inequalities applied to fixed-charge network problems. *Expert Systems with Applications*, 38(6):6627–6636.

Saharidis, G. K., Minoux, M., and Ierapetritou, M. G. (2010). Accelerating benders method using covering cut bundle generation. *International Transactions in Operational Research*, 17(2):221–237.

Santoso, T., Ahmed, S., Goetschalckx, M., and Shapiro, A. (2005). A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research*, 167(1):96–115.

Sherali, H. D. and Lunday, B. J. (2013). On generating maximal nondominated benders cuts. *Annals of Operations Research*, 210:57–72.

Stursberg, P. M. (2019). *On the mathematics of energy system optimization*. PhD thesis, Technische Universität München.

Van Roy, T. J. (1986). A cross decomposition algorithm for capacitated facility location. *Operations Research*, 34(1):145–163.

Wentges, P. (1996). Accelerating benders' decomposition for the capacitated facility location problem. *Mathematical Methods of Operations Research*, 44(2):267–290.

Wu, L. and Shahidehpour, M. (2010). Accelerating the benders decomposition for network-constrained unit commitment problems. *Energy Systems*, 1(3):339–376.

You, F. and Grossmann, I. E. (2013). Multicut benders decomposition algorithm for process supply chain planning under uncertainty. *Annals of Operations Research*, 210:191–211.