Gutenberg School of Management and Economics

& Research Unit “Interdisciplinary Public Policy”

Discussion Paper Series

Benders Decomposition using Core-Maximal Cuts and Its Application to the Single-Picker Routing Problem with Scattered Storage

André Hessenius, Christian Manz, Stefan Irnich

18th December 2025

Discussion paper number 2511

Johannes Gutenberg University Mainz
Gutenberg School of Management and Economics
Jakob-Welder-Weg 9
55128 Mainz
Germany
https://wiwi.uni-mainz.de/

Contact details

André Hessenius
Chair of Logistics Management
Johannes-Gutenberg University
Jakob-Welder-Weg 9
55128 Mainz
Germany
hessenius@uni-mainz.de

Christian Manz
Chair of Logistics and Management
Johannes-Gutenberg University
Jakob-Welder-Weg 9
55128 Mainz
Germany
chmanz@uni-mainz.de

Stefan Irnich
Chair of Logistics and Management
Johannes-Gutenberg University
Jakob-Welder-Weg 9
55128 Mainz
Germany
irnich@uni-mainz.de

Benders Decomposition using Core-Maximal Cuts and Its Application to the Single-Picker Routing Problem with Scattered Storage

André Hessenius[a,*], Christian Manz[a], Stefan Irnich[a]

[a] *Chair of Logistics Management, Department of Business & Economics, Johannes Gutenberg University Mainz, Jakob-Welder-Weg 9, D-55128 Mainz, Germany.*

Abstract

We present a new solution approach to the single-picker routing problem with scattered storage (SPRP-SS) for order picking in warehouses. Heßler and Irnich (INFORMS J. on Comp. 36(6):1417–1435, 2024) demonstrated how to construct an extension to Ratliff and Rosenthal's state space to incorporate the scattered storage dimension. Based on this extended state space, we propose a new formulation that makes two types of decisions: selection of pick positions and determination of the tour that visits the chosen pick positions. This novel formulation is an ideal starting point for a Benders decomposition. For this approach, a fast convergence is crucial. To this end, we analyze properties of the optimality cuts. We examine the most popular type of cut from the literature—the Magnanti-Wong cut—and determine how modified subproblems, such as Magnanti and Wong's, generate stronger cuts than arbitrary subproblem solutions. As a result, we propose core-maximal cuts as a new type of optimality cut. In our computational experiments, we compare the resulting Benders decomposition algorithm with three mixed-integer linear programming formulations, as well as alternative Benders approaches. Core-maximal cuts can be used also for other Benders subproblems as well. We showcase the application to the capacitated facility location problem and the fixed charge network flow problem. Results show that this approach often requires fewer Benders iterations and has shorter total computation times than established Benders decomposition algorithms.

Keywords: Benders decomposition; Benders cuts; routing; warehousing; picker routing; scattered storage

1. Introduction

Activities in a warehouse involve receiving, storing, picking, packing, and shipping goods (Gu et al., 2007). Besides operations such as storage assignment, zoning and batching, routing remains to be one of the main warehouse activities (Boysen et al., 2019; van Gils et al., 2018). In this work, we focus on picking operations in warehouses where pickers move through the warehouse in order to collect articles from the storage locations in response to requests from customers (Masae et al., 2020). This represents the predominant setting in Western Europe, where more than 80% of warehouses use such a system (de Koster et al., 2007). According to de Koster et al. (2007), the picking process is usually the most expensive operation

*Corresponding author.
Email addresses: hessenius@uni-mainz.de (André Hessenius), chmanz@uni-mainz.de (Christian Manz), irnich@uni-mainz.de (Stefan Irnich)

in a warehouse reaching up to 55% of total warehouse operations cost (Drury, 1988; Tompkins et al., 2003; Frazelle, 2002).

The basic *Single-Picker Routing Problem* (SPRP) seeks a minimum-length picker tour to collect a given set of articles from the corresponding positions, where the distances result from a single-block parallel-aisle warehouse layout. The SPRP can be solved very efficiently, since the dynamic programming algorithm of Ratliff and Rosenthal (1983) computes the shortest tour in linear time (Heßler and Irnich, 2022).

To reduce the length of the picker tour, e-commerce companies such as Amazon and Zalando use *scattered-storage* or *mixed-shelves* warehouses. This means that one or more articles can each be picked from different positions. Selecting positions that complement each other well results in shorter picker tours. On the downside, the *Single-Picker Routing Problem with Scattered Storage* (SPRP-SS) has been shown to be NP-hard (Weidinger, 2018), even if routing is restricted to simple routing policies (Lüke et al., 2024).

We briefly formalize the SPRP-SS: For a set S of articles to be collected, each $s \in S$ has a given demand q_s. Additionally, each article is located at possibly multiple positions P_s in the warehouse. The set of all positions is given by $P = \bigcup_{s \in S} P_s$. Each position for each article comes with a supply b_{sp} that gives the maximum amount of article $s \in S$ that the picker can collect at position $p \in P_s$. In case of $q_s = 1$ and $b_{sp} = 1$ for all $s \in S, p \in P_s$, or equivalently $q_s \leq b_{sp}$ for all $s \in S, p \in P_s$, we speak of the *unit-demand* SPRP-SS. In the unit-demand SPRP-SS, the picker only needs to visit one position per article. In the *general-demand* SPRP-SS, however, the picker may have to visit multiple positions to fulfill the demand of each article.

1.1. Extended State Space Formulation

To solve the basic SPRP, Ratliff and Rosenthal (1983) introduce a state space that models the intermediate solutions of a picker tour. This state space is a directed, acyclic graph with non-negative weights for all edges. Each vertex represents a possible state of the picking tour in a given aisle, and each edge represents the options available to the picker in that aisle or between aisles. Ratliff and Rosenthal demonstrate that every path in the state space corresponds to a feasible tour in the warehouse, and vice versa. Thus, searching for the shortest tour is equivalent to searching for the shortest path in the state space, which can be done in linear time (Heßler and Irnich, 2022).

With scattered storage, it is unclear from which position or positions an article should be collected. Therefore, the state space of Ratliff and Rosenthal is insufficient for modeling the SPRP-SS because it does not account for position selection. Heßler and Irnich (2024) introduce an extended state space for this purpose. This extension enables the picker to skip positions and collect the articles from another aisle. However, the equivalence of a path in the state space to a feasible tour in the warehouse no longer holds. While every path in the state space leads to a tour in the warehouse, the problem lies in demand coverage. The picker could skip positions regarding an article and thus not collect a sufficient quantity. To address this issue, Heßler and Irnich present a network-flow formulation to determine a shortest path that fulfills the demand-coverage constraints.

Our work is also based on this idea: we search for the shortest path in the extended state space while ensuring demand is covered. However, we split position selection and routing decisions in the new *position-based* (PB) formulation and use Benders decomposition algorithm to solve it optimally.

1.2. Contributions

The contributions of the paper can be summarized as follows:

1. We present a new *mixed-integer (linear) programming* (MIP) model for the SPRP-SS, the PB formulation, relying on the extended state space proposed by Heßler and Irnich (2024). This type of formulation can be applied to all variants of the SPRP-SS for which a dynamic programming approach is known, including those with different warehouse layouts, different input/output points, unit demand or general demand, etc.
2. We use Benders decomposition to solve the PB formulation. We examine the resulting subproblem and its dual, which generate optimality cuts for the Benders master problem. We provide descriptive insights on why and how the procedure of Magnanti and Wong (1981) works to obtain stronger cuts for the SPRP-SS.
3. We analyze the quality of different types of cuts and establish the conditions for defining a new type of cut: *core-maximal cuts*. We introduce a new subproblem that computes these cuts. This results in fewer iterations and a faster runtime of the Benders algorithm than the subproblems proposed in the existing literature.
4. Our Benders approach usually outperforms (with respect to runtime) the state-of-the-art approach by Heßler and Irnich (2024); Lüke et al. (2025) in unit-demand and general-demand settings.

1.3. Structure

This paper is structured as follows: Section 2 provides an overview of related work on the SPRP-SS and Benders decomposition, focusing on cut enhancement. Section 3 revisits Benders decomposition, introduces notation, presents the new PB formulation for the SPRP-SS, and applies Benders decomposition to the model. Specifically, we examine solving the primal and dual subproblems, provide insights into enhancing optimality cuts using Magnanti and Wong's subproblem, and emphasize the properties of these optimality cuts. Section 4 then introduces the new core-maximal cuts and shows how to generate them by solving a new type of subproblem based. Section 5 presents and discusses computational results, and Section 6 draws final conclusions.

2. Related Work

Most approaches to solving the SPRP-SS rely on mixed-integer linear programming techniques. First approaches involve a reduction to the *traveling purchaser problem* (Singh and van Oudheusden, 1997) as well as a formulation based on the *traveling salesman problem* (TSP) (Daniels et al., 1998). Years later, Weidinger (2018) extends the latter idea with additional constraints and compares it to a novel two-stage decomposition procedure for the SPRP-SS. In contrast, Goeke and Schneider (2021) use a formulation that is inspired by the actions of the state space of Ratliff and Rosenthal (1983) for single-block parallel-aisle warehouses without scattered storage. They model picker movements within aisles and in cross-aisles with binary variables and enforce demand coverage with additional constraints. Heßler and Irnich (2024) use the state spaces of Ratliff and Rosenthal (1983) for single-block parallel-aisle warehouses and the state space of Roodbergen and de Koster (2001) for the two-block warehouses. They incorporate scattered storage by adding additional actions as parallel edges to the network representing the state space. The SPRP-SS is then solved as a MIP, which is a shortest-path problem with demand-coverage constraints. Section 3.2 provides further details.

While previous methods aim to solve the SPRP-SS directly, a decomposition approach could yield an alternative way for solving it exactly. The two most common techniques are Dantzig-Wolfe decomposition, which decomposes a problem based on constraints, and Benders decomposition, which decomposes based on variables (Benders, 1962; Dantzig and Wolfe, 1960). Dantzig-Wolfe decomposition requires column-generation process, whereas Benders decomposition starts with a relaxation of the master problem, adding constraints (cuts) to the relaxation until a provable optimal solution is found. However, neither approach is directly applicable to the state-of-the-art formulation provided by Heßler and Irnich (2024) and Lüke et al. (2025). We will therefore reformulate this model using two types of variables: one for position selection and one for routing decisions. Consequently, a Benders decomposition can be applied for the new formulation.

As Rahmaniani et al. (2017) point out, the classical Benders algorithm can converge slowly, requiring excessive computation time. The quality of the computed cuts is of major importance. Magnanti and Wong (1981) have introduced the concept of a *Pareto-optimal* cut, which is defined as one for which no other cut provides a strictly better bound for all possible master solutions. However, the process of finding Pareto-optimal cuts requires the determination of a point in the relative interior of the integer polyhedron and solving the basic Benders subproblem before the associated *Magnanti-Wong* (MW) subproblem. Various other criteria for good cuts have been proposed. Sherali and Lunday (2013) use *maximal nondominated cuts* to achieve strong bounds while avoiding the additional effort of the MW procedure. They slightly relax the right-hand side of the primal subproblem to achieve this, which results in a perturbation of the coefficients in the dual subproblem's objective. These approaches introduce a direct measure of cut quality. As an alternative approach, Fischetti et al. (2010) search for a minimal set of constraints that render the subproblem infeasible. They compute this infeasibility set from a reformulation of the subproblem as a pure feasibility problem to which they add a normalization constraint that binds the optimal solution, resulting in so-called *minimum infeasible subsystem* (MIS) cuts. Stursberg (2019) builds on this idea, defining the weights in the normalization constraint so that the subproblem results in a facet-defining cut. As with the MW procedure, a point in the relative interior of the master problem polyhedron must be determined. Another common technique is to add multiple cuts per iteration. Saharidis et al. (2010) define a *covering cut bundle* as a set of cuts that provides at least a certain amount of information about each master problem variable. They generate these cuts by repeatedly solving the dual subproblem and changing the bounds of each variable in every iteration to match predefined ranges. In a very recent publication, Glomb et al. (2025) propose a new cut selection strategy, which implements the idea that a good cut should exclude a large set of points from being optimal in the Benders master problem. Their *optimal line-shifting* (OLS) cuts are computed to achieve this goal. The computation is however involved, requiring the solution of a subproblem with a modified feasible region and objective in each iteration. If the incumbent changes, a second subproblem must be solved. According to Glomb et al., Benders algorithms that combine OLS with MIS cuts have been identified to be "faster [than other Benders algorithms] in situations with limited memory or in cases where the subproblem is challenging to solve".

While the aforementioned methods focus on generating strong cuts, Raidl (2015) uses metaheuristics to quickly generate high-quality master solutions and add the corresponding cuts into the Benders master problem. This reduces iterations and overall runtimes. Lin and Üster (2013) propose a similar approach of prematurely stopping the master problem in the early stages to avoid time-consuming iterations. To reduce computation time, Fischetti et al. (2016) and Randazzo et al. (2001) use problem-specific algorithms to solve the subproblem rather than using a general-purpose *linear program(ming)* (LP) solver. Fischetti et al. also propose the generation of initial cuts (for fractional master solutions) before the actual branch-and-cut procedure starts. For this approach, stabilization methods are crucial.

3. Benders Decomposition for the SPRP-SS

Building on the classical Benders decomposition algorithm presented in Section 3.1, we will introduce the PB formulation of the SPRP-SS in Section 3.2. The main content is provided in Section 3.3, where the SPRP-SS-specific Benders decomposition algorithm is described, focusing on subproblem solution and cut enhancement.

3.1. Notation and Classical Benders Decomposition

For a general introduction to Benders decomposition, we refer to (Rahmaniani et al., 2017). The classical version of Benders decomposition can be described by the following type of problem:

$$z_{\text{MIP}} = \min \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y} \tag{1a}$$

$$\text{subject to} \quad A\boldsymbol{x} + F\boldsymbol{y} \geq \boldsymbol{b} \tag{1b}$$

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

with A, F, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}$ of suitable dimension and with integer-valued coefficients, and where all rows of A have at least one non-zero entry. Furthermore, let Y denote the set of feasible solutions that respect all constraints defined by $\boldsymbol{y}$ variables only. The decomposition of problem (1), as suggested by Benders (1962), decomposes the model into a master problem and a subproblem. However, it requires variations of these smaller problems to be solved iteratively. The master problem includes the so-called complicating variables $\boldsymbol{y} \in Y$ and all the constraints defined by these variables alone. In addition, an auxiliary variable z is introduced to provide a lower bound on $\boldsymbol{c}^\top \boldsymbol{x}$ depending on $\boldsymbol{y} \in Y$. We assume that an appropriate lower bound z_0 is chosen for z in order to prevent the master problem becoming unbounded. Additionally, linear constraints, known as *cuts*, are generated iteratively from the dual solution of the subproblem and added to model (2), as stated below. For a given solution $\bar{\boldsymbol{y}} \in Y$ of this *Benders master problem* (BM), the associated *Benders subproblem* (SP($\bar{\boldsymbol{y}}$)), given by model (3), describes the remaining linear optimization problem in the variables $\boldsymbol{x}$.

Master problem BM:

$$\min \quad z + \boldsymbol{d}^\top \boldsymbol{y} \tag{2a}$$

$$\text{s.t.} \quad \{cuts\} \tag{2b}$$

$$\boldsymbol{y} \in Y, z \geq z_0 \tag{2c}$$

Subproblem SP($\bar{\boldsymbol{y}}$):

$$\min \quad \boldsymbol{c}^\top \boldsymbol{x} \tag{3a}$$

$$\text{s.t.} \quad A\boldsymbol{x} \geq \boldsymbol{b} - F\bar{\boldsymbol{y}} \tag{3b}$$

$$\boldsymbol{x} \geq \boldsymbol{0} \tag{3c}$$

Dual subproblem DSP($\bar{\boldsymbol{y}}$):

$$\max \quad \boldsymbol{\pi}^\top (\boldsymbol{b} - F\bar{\boldsymbol{y}}) \tag{4a}$$

$$\text{s.t.} \quad A^\top \boldsymbol{\pi} \leq \boldsymbol{c} \tag{4b}$$

$$\boldsymbol{\pi} \geq \boldsymbol{0} \tag{4c}$$

Benders' algorithm starts by solving BM with an empty set of cuts in (2b). In each iteration, the optimal solution $\bar{\boldsymbol{y}}$ of the Benders master is passed to the subproblem. More precisely, what needs to be computed is a dual optimal solution $\bar{\boldsymbol{\pi}}$ of the subproblem, which can be obtained from either the primal subproblem $\mathrm{SP}(\bar{\boldsymbol{y}})$ or directly as a solution to its dual problem $\mathrm{DSP}(\bar{\boldsymbol{y}})$, given by model (4). Notice that the feasible region of the dual subproblem $\mathrm{DSP}(\bar{\boldsymbol{y}})$ remains unchanged throughout the entire process. Consequently, any dual feasible solution obtained in an arbitrary iteration is a feasible dual solution for each and every solution $\bar{\boldsymbol{y}}$ of BM.

When solving the dual subproblem $\mathrm{DSP}(\bar{\boldsymbol{y}})$ for a given master solution $\bar{\boldsymbol{y}} \in Y$, three cases can arise:

1. Infeasible: If the dual subproblem $\mathrm{DSP}(\bar{\boldsymbol{y}})$ is infeasible, the primal subproblem is either unbounded or infeasible. Since the dual feasible region is independent of $\bar{\boldsymbol{y}} \in Y$, the given problem (1) is then also infeasible. For the following, we assume that problem (1) has a feasible solution.
2. Unbounded: If the dual subproblem $\mathrm{DSP}(\bar{\boldsymbol{y}})$ is unbounded, the primal subproblem $\mathrm{SP}(\bar{\boldsymbol{y}})$ must be infeasible. Let $\bar{\boldsymbol{\pi}}$ be an extreme ray of the dual subproblem improving the objective, i.e., $\bar{\boldsymbol{\pi}}^\top(\boldsymbol{b}-F\bar{\boldsymbol{y}}) > 0$. To prevent $\bar{\boldsymbol{y}}$ as a master solution, a *feasibility cut* of the form $\bar{\boldsymbol{\pi}}^\top(\boldsymbol{b}-F\boldsymbol{y}) \leq 0$ can be added to (2b).
3. Bounded and feasible: If the dual subproblem $\mathrm{DSP}(\bar{\boldsymbol{y}})$ has an optimal solution $\bar{\boldsymbol{\pi}}$, also the primal subproblem $\mathrm{SP}(\bar{\boldsymbol{y}})$ has one, say $\bar{\boldsymbol{x}}$, and the optimal objective values are identical, i.e., $\boldsymbol{c}^\top\bar{\boldsymbol{x}} = \bar{\boldsymbol{\pi}}^\top(\boldsymbol{b} - F\boldsymbol{y})$, due to strong duality. Then,
$$\bar{\boldsymbol{\pi}}^\top(\boldsymbol{b} - F\boldsymbol{y}) \leq z \tag{5}$$
is a valid inequality for all possible $\boldsymbol{y} \in Y$, since the dual feasible region does not change such that any feasible solution to the dual subproblem provides a lower bound on the objective value of the primal subproblem due to weak duality. If (5) is violated for the master solution $\bar{\boldsymbol{y}}$, it can be added to (2b) as a so-called *optimality cut*. Otherwise, an optimal solution $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$ to problem (1) has been found, and Benders' algorithm terminates.

3.2. Position-based Formulation of the SPRP-SS

In this section, we modify the binary formulation of Heßler and Irnich (2024) for the SPRP-SS to enable the application of Benders' algorithm. We refer to their model as the HI formulation. It is a shortest-path model extended by demand-covering constraints. The exact formalization of the construction of the underlying network defined over a digraph $G = (V, E)$ is beyond the scope of this paper. It is sufficient to know that the digraph is constructed from the dynamic-programming state space of the associated *dynamic program* (DP) for the standard SPRP (without scattered articles). For single-block rectangular warehouses, this is the well-known DP of Ratliff and Rosenthal (1983). For two-block rectangular warehouses, Roodbergen and de Koster (2001) provide an adapted DP. For other warehouse layouts and I/O point configurations, we refer to the overview provided in (Heßler and Irnich, 2024, Table 1). These DPs have in common that they decide, aisle by aisle or stage by stage, how the picker will traverse a cross-aisle or an aisle. More precisely, aisles might be divided into sub-aisles (e.g., two sub-aisles per aisle in a two-block warehouse). To simplify the description, we will only refer to aisles. Summarizing, let $E = \bigcup_{j\in J}(E_j^{aisle} \cup E_j^{cross})$ denote the edges, indexed by aisle $j \in J$, that describe aisle actions and cross-aisle actions, respectively.

Heßler and Irnich (2024) showed how to extend these state spaces to incorporate the scattered storage dimension. This extension involves adding extra parallel edges to E_j^{aisle} for each aisle $j \in J$ where scattering allows to skip some positions $p \in P$ while still being able to collect the required quantity q_s of all articles $s \in S$ from alternative aisles.

For the formalization of the new PB model, let $\mathcal{N}$ be the vertex-arc incidence matrix of the digraph G, let $o \in V$ denote the initial and $d \in V$ the final vertex, and let $\boldsymbol{u}_o$ and $\boldsymbol{u}_d$ be the corresponding unit vectors. The variables $\boldsymbol{x} = (x_e)_{e \in E}$ in the HI formulation indicate the flow through each edge $e \in E$. The shortest-path sub-model of HI is $\{\min \boldsymbol{c}^\top \boldsymbol{x} \text{ subject to } \mathcal{N}\boldsymbol{x} = \boldsymbol{u}_o - \boldsymbol{u}_d, \boldsymbol{x} \geq \boldsymbol{0}\}$. Clearly, this type of model contains only one type of variable, and is, hence, not suitable for a Benders decomposition. Therefore, we introduce additional variables $\boldsymbol{y} = (y_p)_{p \in P} \in \{0,1\}^P$ that indicate whether the picker has to visit pick position $p \in P$. Recall from Section 1 that b_{sp} denotes the number of *stock keeping units* (SKUs) of article $s \in S$ that can be collected from pick position $p \in P$. The demand-covering constraints can then be stated as $\sum_{p \in P_s} b_{sp} y_p \geq q_s$ for all $s \in S$. Moreover, the *conflict set* $\mathcal{C} \subset E \times P$ contains those pairs (e, p) of edges and positions that refer to the same aisle in which aisle-action e omits position p, i.e., does *not* allow to collect SKUs from position p. The new PB formulation of the SPRP-SS is a MIP model that reads as follows:

$$z_{\text{SPRP-SS}} = \min \quad \boldsymbol{c}^\top \boldsymbol{x} \tag{6a}$$

$$\text{subject to} \quad \mathcal{N}\boldsymbol{x} = \boldsymbol{u}_o - \boldsymbol{u}_d \tag{6b}$$

$$\sum_{p \in P_s} b_{sp} y_p \geq q_s \qquad \forall s \in S \tag{6c}$$

$$\sum_{(e,p) \in \mathcal{C}} x_e + y_p \leq 1 \qquad \forall p \in P \tag{6d}$$

$$\boldsymbol{x} \geq \boldsymbol{0}, \qquad \boldsymbol{y} \in \{0,1\}^P \tag{6e}$$

The objective (6a) minimizes the total routing cost, while constraints (6b) model flow conservation. Constraints (6c) are the demand-covering constraints. Conflict constraints (6d) state that if SKUs are to be collected from a position, none of the edges that are in conflict with that position can be used in the solution. The domains of the variables are specified by (6e).

3.3. Benders Decomposition of the Position-based Formulation

Formulation (6) is an ideal starting point for a Benders decomposition. The position variables $\boldsymbol{y} = (y_p)_{p \in P}$ are the complicating variables, which allows us to define $Y = \{\boldsymbol{y} \in \{0,1\}^P : \sum_{p \in P_s} b_{sp} y_p \geq q_s \text{ for all } s \in S\}$. For each solution $\bar{\boldsymbol{y}} \in Y$ of BM, the subproblem $\text{SP}(\bar{\boldsymbol{y}})$ comprises flow variables $\boldsymbol{x} = (x_e)_{e \in E}$ defining a modified shortest-path model:

$$z_{\text{BM}} = \min \quad z \tag{7a}$$

$$\text{subject to} \quad \{\text{optimality cuts}\} \tag{7b}$$

$$\boldsymbol{y} \in Y, z \geq 0 \tag{7c}$$

$$z_{\text{SP}(\bar{\boldsymbol{y}})} = \min \quad \boldsymbol{c}^\top \boldsymbol{x} \tag{8a}$$

$$\text{subject to} \quad \mathcal{N}\boldsymbol{x} = \boldsymbol{u}_o - \boldsymbol{u}_d \tag{8b}$$

$$\sum_{(e,p) \in \mathcal{C}} x_e \leq 1 - \bar{y}_p \qquad \forall p \in P \tag{8c}$$

$$\boldsymbol{x} \geq \boldsymbol{0} \tag{8d}$$

The interpretation is rather straightforward: The Benders master problem (7) chooses a feasible combination of pick positions (feasibility is incorporated into the definition of the set Y). This implies that only optimality cuts are needed in (7b) for the SPRP-SS. The Benders subproblem (8) is still a shortest o-d-path model where constraints (8c) forbid and eliminate any flow over an edge in conflict with the chosen master solution $\bar{\boldsymbol{y}}$, i.e., the positions $P^1(\bar{\boldsymbol{y}}) = \{p \in P : \bar{y}_p = 1\}$.

For the following subsections, we assume that a feasible master solution $\bar{\boldsymbol{y}} \in Y$ is given and fixed.

3.3.1. Solution of the Primal Subproblem

There is no need to use an LP solver to compute an optimal solution to the primal subproblem $\mathrm{SP}(\bar{\boldsymbol{y}})$. For each chosen position, i.e., $p \in P^1(\bar{\boldsymbol{y}})$, we can eliminate the conflicting edges $e \in E$, i.e., those with $(e,p) \in \mathcal{C}$ from the digraph G. Let $G(\bar{\boldsymbol{y}}) = (V, E(\bar{\boldsymbol{y}}))$ denote the resulting digraph. A shortest o-d-path in $G(\bar{\boldsymbol{y}})$ can be computed with an appropriate labeling algorithm. We can exploit that G and herewith also $G(\bar{\boldsymbol{y}})$ is acyclic and the stages of the DP give rise to a natural topological sorting. The shortest path defines values for $\bar{\boldsymbol{x}}$ as well as distance labels $L(v)$ describing the length of a shortest path from o to v for all vertices $v \in V$: Set $\bar{x}_e = 1$ for edges belonging to the shortest o-d-path, and set $\bar{x}_e = 0$ for all other edges e. The distance labels $L(v)$ fulfill $L(d) - L(o) = z_{\mathrm{SP}(\bar{\boldsymbol{y}})}$. They are useful for the solution of the dual subproblem $\mathrm{SP}(\bar{\boldsymbol{y}})$, which we analyze next.

3.3.2. Dual Subproblem

While the primal subproblem $\mathrm{SP}(\bar{\boldsymbol{y}})$ can be solved over a reduced digraph consisting of only non-conflicting edges, the dual subproblem $\mathrm{DSP}(\bar{\boldsymbol{y}})$ must not be considered without these conflicts. We will see that the values of the dual variables to the conflict constraints (8c) have a pivotal role in the determination of effective optimality cuts. Formally, let $\boldsymbol{\mu} = (\mu_v)_{v \in V} \in \mathbb{R}^V$ be the dual variables to (8b) and $\boldsymbol{\nu} = (\nu_p)_{p \in P} \leq \mathbf{0}$ to (8c). The dual subproblem $\mathrm{DSP}(\bar{\boldsymbol{y}})$ is of the following form:

$$z_{\mathrm{DSP}(\bar{\boldsymbol{y}})} = \max \quad \mu_o - \mu_d + \sum_{p \in P} \nu_p (1 - \bar{y}_p) \tag{9a}$$

$$\text{subject to} \quad \mu_i - \mu_j + \sum_{(e,p) \in \mathcal{C}} \nu_p \leq c_e \qquad \forall e \in E, e = (i,j) \tag{9b}$$

$$\boldsymbol{\mu} \in \mathbb{R}^V, \quad \boldsymbol{\nu} \leq \mathbf{0} \tag{9c}$$

For a feasible dual solution $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}})$, the optimality cut (see Eq. (5) for the general case) is here

$$z \geq \bar{\mu}_o - \bar{\mu}_d + \sum_{p \in P} \bar{\nu}_p (1 - y_p) \tag{10a}$$

where the right-hand side of the optimality cut has the constant part and variable part

$$\mathrm{const}(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}}) := \bar{\mu}_o - \bar{\mu}_d + \sum_{p \in P} \bar{\nu}_p \qquad \text{and} \qquad \mathrm{var}(\bar{\boldsymbol{\nu}}) := -\sum_{p \in P} \bar{\nu}_p y_p, \tag{10b}$$

respectively. For analyzing the structure of the subproblems, it is convenient to partition the set P of all positions into $P^0(\bar{\boldsymbol{y}}) = \{p \in P : \bar{y}_p = 0\}$ and $P^1(\bar{\boldsymbol{y}}) = \{p \in P : \bar{y}_p = 1\}$. Accordingly, we can arrange the dual variables referring to the conflict constraints into $\boldsymbol{\nu} = (\boldsymbol{\nu}^0, \boldsymbol{\nu}^1)$. In the objective (9a), only the variables $\boldsymbol{\nu}^0$ occur with objective coefficient equal to 1, while the variables $\boldsymbol{\nu}^1$ are not present. Since all $\boldsymbol{\nu}$-variables are non-positive, see (9c), the variables $\boldsymbol{\nu}^1$ can be set to any non-positive value that is beneficial. In particular, if a $\boldsymbol{\nu}^1$-variable occurs in a constraint (9b), it can be set to any value $-M$ (for a large number $M \gg 0$) so that the constraint is fulfilled.

3.3.3. Solution of the Dual Subproblem

With the above analysis, we are now able to describe an optimal dual solution.

Proposition 1. *(a) The solution* $(\boldsymbol{\mu}, \boldsymbol{\nu}) = (\boldsymbol{\mu}, (\boldsymbol{\nu}^0, \boldsymbol{\nu}^1))$ *defined by*

$$\mu_v = -L(v), \forall v \in V, \qquad \boldsymbol{\nu}^0 = \mathbf{0}, \qquad \textit{and} \qquad \boldsymbol{\nu}^1 = (-M),$$

is an optimal solution to the dual subproblem DSP($\bar{\boldsymbol{y}}$)*, where* $L(v), v \in V$ *are the shortest-path distance labels computed over* $G(\bar{\boldsymbol{y}})$ *(see Section 3.3.1), and* M *is a large constant. The computation of the solution can be performed in linear time w.r.t. the size of the underlying state space* G*.*
(b) The parameter can be chosen as $M = -R$*, with* R *denoting the length of the shortest tour that collects SKUs from all positions* $p \in P$*.*
(c) For two dual optimal solutions $(\boldsymbol{\mu}', \boldsymbol{\nu}')$ *and* $(\boldsymbol{\mu}'', \boldsymbol{\nu}'')$*, the total variation in the dual prices of the variables* $\boldsymbol{\nu}$ *can be assumed upper bounded by* $|P| \cdot R$*.*

Proof. Proof (a) We first show that $(\boldsymbol{\mu}, \boldsymbol{\nu})$ is a feasible solution to (9): For each edge $e = (i, j) \in E(\bar{\boldsymbol{y}})$, by construction, there is no conflict to any position $p \in P^1(\bar{\boldsymbol{y}})$ such that the sum $\sum_{(e,p)\in\mathcal{C}} \nu_p$ in (9b) is zero. Hence, the inequality (9b) reduces to $\mu_i - \mu_j \le c_e$, which is the label optimality condition fulfilled for all $e = (i, j) \in E(\bar{\boldsymbol{y}})$ (see Ahuja et al., 1993, Section 5.2). For each edge $e = (i, j) \in E \setminus E(\bar{\boldsymbol{y}})$, there exists at least one conflicting position $p \in P^1(\bar{\boldsymbol{y}})$. As a result, the sum $\sum_{(e,p)\in\mathcal{C}} \bar{\nu}_p$ in (9b) evaluates to a value $\le -M$ such that (9b) is fulfilled. Together, feasibility of $(\boldsymbol{\mu}, \boldsymbol{\nu})$ is shown.

The objective value in (9a) is $\mu_o - \mu_d + \sum_{p\in P} \nu_p(1 - \bar{y}_p) = -L(o) + L(d) = z_{\mathrm{SP}(\bar{\boldsymbol{y}})}$. It follows from weak duality that $(\boldsymbol{\mu}, \boldsymbol{\nu})$ is an optimal solution.
(b) Let $p \in P$ be any position. The dual variable ν_p appears in all constraints (9b) for which the edge $e = (i, j) \in E$ is in conflict to p, i.e., $(e, p) \in \mathcal{C}$. For such an edge $e = (i, j)$, the difference of the μ-variables in constraint $\mu_i - \mu_j + \nu_p \le c_{ij}$ is inherently bounded by the length of the shortest tour that collects all SKUs, denoted by R. Therefore, choosing $\nu_p \le -R$ is always sufficient to fulfill any constraint (9b). Note that the shortest tour collecting all items can be estimated by the length of a tour that follows the S-Shape strategy (see Petersen (1997)).
(c) According to (b), dual optimal solutions can be assumed to have $\boldsymbol{\nu}$-variables with value limited to the interval $[-R, 0]$. For two such dual optimal solutions, $\boldsymbol{\nu}'$ and $\boldsymbol{\nu}''$, the total variation in the dual prices, i.e., $\sum_{p\in P} |\nu'_p - \nu''_p|$ is therefore upper bounded by $|P| \cdot R$. □

The dual subproblem has infinitely many feasible solutions and, especially, also infinitely many optimal solutions. Since the $\boldsymbol{\mu}$-variables always appear as differences, one can add an arbitrary constant to the values of these variables without compromising optimality. Furthermore, lowering the value of any variable ν_p for $p \in P^1(\bar{\boldsymbol{y}})$ (making its absolute value larger) is an operation that preserves optimality. Formally, if $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}})$ is optimal, then, for any value $q \in \mathbb{R}$, also the dual solution $(\bar{\boldsymbol{\mu}}', \bar{\boldsymbol{\nu}}')$ is optimal for $\bar{\mu}'_v = \bar{\mu}_v + q$ for all $v \in V$, $\bar{\boldsymbol{\nu}}'^0 = \bar{\boldsymbol{\nu}}^0$, and $\bar{\boldsymbol{\nu}}'^1 \le \bar{\boldsymbol{\nu}}^1$.

Example 1. *Figure 1 visualizes an SPRP-SS instance defined over a single-block warehouse with three aisles and nine cells per aisle. The pick list comprises seven articles from which only article* $s = 1$ *is scattered and available in cells 0 and 7 in the rightmost aisle. Therefore, only two corresponding position denoted by* p_0 *and* p_7 *are of interest, since every feasible picker tours must visit the other positions. Accordingly,* *we can describe all master solutions with $(\bar{y}_0, \bar{y}_7)$ only. We assume a unit-demand instance, so that the capacity constraint (6c) of article $s = 1$ is $y_0 + y_7 \geq 1$. Consequently, $(1,0)$, $(0,1)$, and $(1,1)$ are the only feasible solutions of BM. All three corresponding, optimal picker tours contain the actions, i.e., picker movements, indicated by the straight green edges in the figure (the cycle $(depot, v_{0,0}, v_{1,0}, v_{1,1}, depot)$). They differ in the part indicated with the dashed lines. Indeed, the master solution $(\bar{y}_0, \bar{y}_7) = (1,0)$ adds the upper movement $(v_{1,0}, v_{2,0}, v_0, v_{2,0}, v_{1,0})$ to the cycle (depicted in blue), $(\bar{y}_0, \bar{y}_7) = (0,1)$ adds the lower movement $(v_{1,1}, v_{2,1}, v_7, v_{2,1}, v_{1,1})$ to the cycle (depicted in orange), and $(\bar{y}_0, \bar{y}_7) = (1,1)$ both parts.*

We focus on the dual solution $(\bar{y}_0, \bar{y}_7) = (0,1)$ (imposing the tour using the dashed movements depicted in orange). Proposition 1 allows us to compute the dual optimal solution, first setting $\bar{\nu}_0 = 0$ and $\bar{\nu}_7 = -R = -50$. In turn, the shortest-path computation must consider these $\boldsymbol{\nu}$-values in the constraint system shown in Figure 1b. The value $\bar{\nu}_0 = 0$ keeps edge $\{2,5\}$ marked with 'bottom (4)' at its original value, while the value $\nu_7 = -50$ relaxes the edge $\{1,4\}$ marked with 'top (2)'. This yields $\bar{\mu}_4 = -48$, $\bar{\mu}_5 = -36$, $\bar{\mu}_6 = -42$, and $\bar{\mu}_d = -36$. The resulting optimality cut is therefore $z \geq \bar{\mu}_o - \bar{\mu}_d + \bar{\nu}_0(1-y_0) + \bar{\nu}_7(1-y_7) = 0 - (-36) + 0(1-y_0) + (-50)(1-y_7)$ which is $z \geq -14 + 50y_7$.

3.4. Magnanti-Wong Cuts

In the conventional Benders approach, the above result of Proposition 1 is sufficient: Iteratively, solve BM, i.e., problem (7), retrieve a solution $\bar{\boldsymbol{y}} \in Y$, solve $\mathrm{SP}(\bar{\boldsymbol{y}})$ or $\mathrm{DSP}(\bar{\boldsymbol{y}})$, and add the optimality cut (9) to BM. However, we observed in pretests that the conventional Benders approach usually converges very slowly.

Magnanti and Wong (1981) defined the concept of Pareto-optimal solutions and cuts to accelerate convergence. In the SPRP-SS context, a dual solution $(\boldsymbol{\mu}', \boldsymbol{\nu}')$ is Pareto-optimal, if no alternative dual optimal solution $(\boldsymbol{\mu}, \boldsymbol{\nu})$ exists such that $\mu_o - \mu_d + \sum_{p \in P} \nu_p(1-y_p) \geq \mu'_o - \mu'_d + \sum_{p \in P} \nu'_p(1-y_p)$ for all $\boldsymbol{y} = (y_p)_{p \in P} \in Y$. In addition, the inequality must be strict for at least one $\boldsymbol{y} \in Y$.

To compute a Pareto-optimal solution, a second linear program must be solved. To this end, let $\tilde{\boldsymbol{y}}$ be an arbitrary point in the relative interior $\mathrm{relint}(Y^c)$ of the convex hull $Y^c = \mathrm{conv}(Y)$. For the SPRP-SS, the second linear program, denoted by $\mathrm{MW}(\bar{\boldsymbol{y}}, \tilde{\boldsymbol{y}})$, is of the following form:

$$z_{\mathrm{MW}(\bar{\boldsymbol{y}},\tilde{\boldsymbol{y}})} = \max \quad \mu_o - \mu_d + \sum_{p \in P} \nu_p(1-\tilde{y}_p) \tag{12a}$$

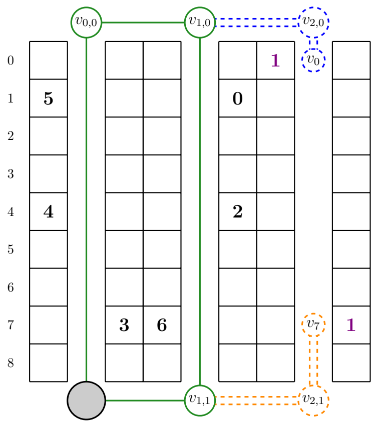
$$\text{subject to} \quad \mu_o - \mu_d + \sum_{p \in P} \nu_p(1-\bar{y}_p) = z_{\mathrm{SP}(\bar{\boldsymbol{y}})} \tag{12b}$$

$$\text{(9b) and (9c)} \tag{12c}$$

Note that $\tilde{y}_p$ in the objective (12a) replaces $\bar{y}_p$ in the objective (9a) of the dual subproblem $\mathrm{DSP}(\bar{\boldsymbol{y}})$. The additional constraint (12b) ensures that the solution remains dual optimal. We will refer to this constraint as the *MW constraint* in the following.

Proposition 2. *For any $\tilde{\boldsymbol{y}} \in relint(Y^c)$, an optimal solution $(\boldsymbol{\mu}^*, \boldsymbol{\nu}^*)$ to the linear program (12) provides an optimality cut of the form (10a) that is Pareto-optimal.*

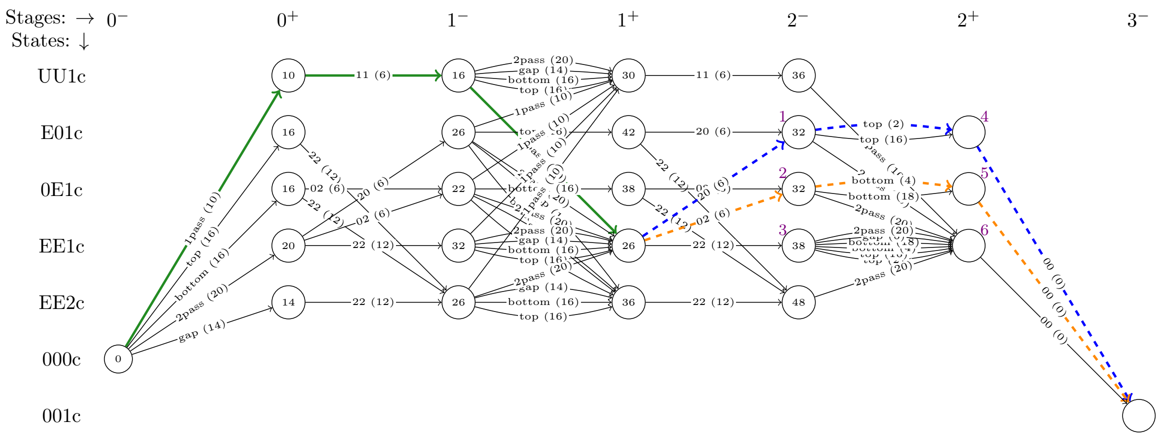
Proof. Proof Magnanti and Wong (1981) show the Pareto-optimality of their cuts in the general context. We additionally provide an SPRP-SS-tailored proof in Appendix A. □



(a) Warehouse with seven articles $S = \{0, 1, \ldots, 6\}$, article $s = 1$ being available at two positions, in cells 0 and 7 in rightmost aisle. The dual prices for these two positions are denoted according to their cell positions, i.e., ν_0 and ν_7.

$$
\begin{aligned}
\max\ & \mu_o - \mu_d + (1 - \bar{y}_0)\nu_0 + (1 - \bar{y}_7)\nu_7 \\
& \vdots \\
& \mu_1 - \mu_4 \qquad + \nu_7 \le 2 \\
& \mu_1 - \mu_4 \qquad \le 16 \\
& \mu_2 - \mu_5 + \nu_0 \qquad \le 4 \\
& \mu_2 - \mu_5 \qquad \le 18 \\
& \cdots \\
& \vdots \\
& \mu_3 - \mu_6 \qquad + \nu_7 \le 2 \\
& \mu_3 - \mu_6 + \nu_0 \qquad \le 4 \\
& \mu_3 - \mu_6 \qquad \le 6
\end{aligned}
$$

(b) Relevant constraints for the implicit solution process. Note that all distance labels up to stage 2^- are independent of $\bar{\boldsymbol{y}} \in Y$. Redundant constraints are omitted also after stage 2^-.



(c) State space (reduced to only the necessary states). Each is marked with its action and routing cost. The distance label of a state v indicates the shortest o-v-path distance, which is independent of $\boldsymbol{y}$ up to stage 2^-. States marked with additional numbers 1 to 6 (in violet) are those relevant for the constraint system shown in Figure 1b, in which dual variables $\boldsymbol{\mu}$ are indexed with these numbers.

Figure 1: SPRP-SS instance with warehouse, some constraints of the dual subproblem that are of interest, and state space.

To generate a point $\tilde{\boldsymbol{y}} \in \text{relint}(Y^c)$ in the relative interior, we use a set of extreme points $\tilde{Y}$ that is generated beforehand as follows: Let $P^{\text{req}} = \bigcap_{y \in Y} P^1(y)$ denote the set of *required* positions, i.e., those that have to be visited in every feasible master solution $\boldsymbol{y} \in Y$. Equivalently, $y_p = 1$ holds for all $\boldsymbol{y} \in Y$ for $p \in P^{\text{req}}$. Conversely, for each $p' \in P \setminus P^{\text{req}}$, there must exist a $\boldsymbol{y} \in Y$ with $y_{p'} = 0$. In particular, the solution $\boldsymbol{y}^{(p')}$ defined by $y_{p'}^{(p')} = 0$ and $y_p^{(p')} = 1$ for all $p \neq p'$ is also a feasible master solution, i.e., $\boldsymbol{y}^{(p')} \in Y$. We define $\tilde{Y} = \bigcup_{p' \in P \setminus P^{\text{req}}} \{\boldsymbol{y}^{(p')}\}$. In every Benders iteration, we sample a new convex combination of the points $\tilde{Y}$ by drawing random weights.

Note that for all $p \in P^{\text{req}}$, the coefficient of ν_p in the MW objective (12a) is zero, such that the value of ν_p can be chosen arbitrarily (non-positive). This is logical consistent, since the Benders master always selects every $p \in P^{\text{req}}$, which does not have to be corrected by an optimality cut.

Solving the MW subproblem (12) for the SPRP-SS is not as simple as solving its standard primal and dual subproblems. The reason is that due to fractional values of $\tilde{\boldsymbol{y}}$, there is a non-trivial interrelation between the variables $\boldsymbol{\mu}$, $\boldsymbol{\nu}^0$, and $\boldsymbol{\nu}^1$. To analyze subproblem (12), we can rewrite the objective (12a) and the constraint (12b) into:

$$z_{\text{MW}(\bar{\boldsymbol{y}},\tilde{\boldsymbol{y}})} = \max \quad \underbrace{\mu_o - \mu_d + \sum_{p \in P^0(\bar{\boldsymbol{y}})} \nu_p}_{= z_{\text{SP}(\bar{\boldsymbol{y}})}} \quad \underbrace{- \sum_{p \in P^0(\bar{\boldsymbol{y}})} \nu_p \tilde{y}_p}_{\text{see Observation 2}} \quad \underbrace{+ \sum_{p \in P^1(\bar{\boldsymbol{y}})} \nu_p (1 - \tilde{y}_p)}_{\text{see Observation 1}} \tag{13a}$$

$$\text{subject to} \quad \mu_o - \mu_d + \sum_{p \in P^0(\bar{\boldsymbol{y}})} \nu_p = z_{\text{SP}(\bar{\boldsymbol{y}})} \tag{13b}$$

We start by analyzing the impact of the $\boldsymbol{\nu}^1$-variables, i.e., ν_p for $p \in P^1(\bar{\boldsymbol{y}})$. Recall that for $p \in P^{\text{req}} \cap P^1(\bar{\boldsymbol{y}})$, the value of ν_p can be set to any non-positive value (the coefficient in the objective and the MW constraint (13b) is zero). For all $p \in P^1(\bar{\boldsymbol{y}}) \setminus P^{\text{req}}$, $\nu_p \leq 0$ and $(1 - \tilde{y}_p) > 0$ holds. Improving the MW objective (13a) with these (non-positive) dual variables results from moving them to a larger value, i.e., closer to 0. For the dual optimal solution for $\text{SP}(\bar{\boldsymbol{y}})$ constructed in Proposition 1, it means that instead of using general dual values $-M$, one should use position-specific values as large as possible. In other words, we must set the $\boldsymbol{\nu}^1$-variables to the largest values possible that do not lead to a violation of a constraint (note that the MW constraint (13b) is not at all affected). However, the constraints (9b) have to be considered, limiting the possibility to move these values completely to 0.

Proposition 3. *If $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}})$ provides a Pareto-optimal cut, then all ν_p for $p \in P^1(\bar{\boldsymbol{y}}) \setminus P^{req}$ are maximal, i.e., they cannot be increased [towards 0] without violating constraints* (13b) *or changing some other dual value $\bar{\nu}_p$. This implies that all components of $\bar{\boldsymbol{\nu}}^1$ can be assumed maximal.*

Proof. Proof Assume that $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}})$ provides a Pareto-optimal cut and that for one position $p_1 \in P^1(\bar{\boldsymbol{y}}) \setminus P^{\text{req}}$ the value $\bar{\nu}_{p_1}$ could be increased to $\bar{\nu}_{p_1} + \varepsilon$ (for some $\varepsilon > 0$) without violating (9b) (note that constraints (12b) equivalent to (13b) are obviously fulfilled). Then, also $(\dot{\boldsymbol{\mu}}, \dot{\boldsymbol{\nu}})$ defined by $\dot{\boldsymbol{\mu}} = \bar{\boldsymbol{\mu}}$, $\dot{\nu}_p = \bar{\nu}_p$ for all $p \in P \setminus \{p_1\}$ and $\dot{\nu}_{p_1} = \bar{\nu}_{p_1} + \varepsilon$ is a dual optimal solution. For all $\boldsymbol{y}^\star \in Y$ with $y^\star_{p_1} = 0$, i.e., $(1 - y^\star_{p_1}) = 1$, it follows

$$\begin{aligned} \bar{\mu}_o - \bar{\mu}_d + \sum_{p \in P} \bar{\nu}_p (1 - y^\star_p) &= \dot{\mu}_o - \dot{\mu}_d + \bar{\nu}_{p_1} + \sum_{p \in P \setminus \{p_1\}} \dot{\nu}_p (1 - y^\star_p) \\ &< \dot{\mu}_o - \dot{\mu}_d + \dot{\nu}_{p_1} + \sum_{p \in P \setminus \{p_1\}} \dot{\nu}_p (1 - y^\star_p) = \dot{\mu}_o - \dot{\mu}_d + \sum_{p \in P} \dot{\nu}_p (1 - y^*_p) \end{aligned}$$

Since $p_1 \notin P^{\text{req}}$, there exists such a $\boldsymbol{y}^\star \in Y$ with $y^\star_{p_1} = 0$. For all other $\boldsymbol{y} \in Y$ with $y_{p_1} = 1$, the cut provided by $(\dot{\boldsymbol{\mu}}, \dot{\boldsymbol{\nu}})$ trivially provides the same lower bound on z (the right hand side of the optimality cut (10a)) as $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}})$, i.e., equality between the above terms on the left-hand side and right-hand side. This contradicts with the assumption that $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}})$ provides a Pareto-optimal cut.

Finally, recall that the values of ν_p for $p \in P^{\text{req}}$ can be set to an arbitrary non-positive value. Setting those to 0 shows that all components of $\bar{\boldsymbol{\nu}}^1$ can be assumed maximal. □

Observation 1. *The MW subproblem (12) pushes the (non-positive) dual prices of* $\boldsymbol{\nu}^1 = (\nu_p)_{p \in P^1(\bar{\boldsymbol{y}})}$ *towards* 0. *Note that* $const(\boldsymbol{\mu}, \boldsymbol{\nu}) = z_{SP(\bar{\boldsymbol{y}})} + \sum_{p \in P^1(\bar{\boldsymbol{y}})} \nu_p$ *holds, resulting from (10b) and (13b). Consequently, the MW subproblem tends to provide larger constants than optimality cuts derived by generic dual optimal solutions.*

Example 2. *(continued from Example 1) We can use Observation 1 to improve upon the previous optimality cut* $z \geq -14 + 50y_7$. *Pushing the value of* ν_7 *from* -50 *to* -2 *does not change the objective value. Indeed, this is sufficient to still 'increase' the cost of the edge* $\{1, 4\}$ *marked with 'top (2)' from 2 to only 4, see Figure 1c. However, the resulting labels give* $\bar{\mu}_4 = -36$, $\bar{\mu}_5 = -36$, $\bar{\mu}_6 = -42$, *and* $\bar{\mu}_d = -36$. *This yields a new optimality cut* $z \geq 34 + 2y_7$, *which is strictly dominating the previous cut.*

For a given $\bar{\boldsymbol{\nu}}$, a dual optimal solution $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}})$ to the Benders subproblem (9), to the MW subproblem (12), and one that provides any other Pareto-optimal cut necessarily maximizes $\bar{\mu}_o - \bar{\mu}_d$. This maximization problem can be solved efficiently:

Proposition 4. *Let* $(\boldsymbol{\mu}, \boldsymbol{\nu})$ *be a dual feasible solution. There exists an efficient algorithm that constructs a dual feasible solution* $(\boldsymbol{\mu}^\star, \boldsymbol{\nu})$ *(one with identical values for the* $\boldsymbol{\nu}$*-variables) that maximizes* $\mu^\star_o - \mu^\star_d$.

Proof. Proof Define $\tilde{c}_e = c_e - \sum_{(e,p) \in \mathcal{C}} \nu_p$. Constraints (9b) can then be restated as $\mu_i - \mu_j \leq \tilde{c}_e$ for all $e \in E$. These are path-optimality conditions for the weighted digraph $(V, A, \tilde{c})$. Therefore, maximizing $\mu^\star_o - \mu^\star_d$ is equivalent to solving the corresponding o-d-shortest path problem on $(V, A, \tilde{c})$. As in Section 3.3.1, we can use a labeling algorithm to compute minimum-distance labels $L(v)$ for all $v \in V$, exploiting that the graph is acyclic and the stages of the DP provide a topological ordering of the vertices. To obtain the solution, we set $\mu^\star_v = -L(v)$ for all $v \in V$. □

Next, we will analyze the impact of the $\boldsymbol{\nu}^0$-variables, i.e., the variables ν_p for $p \in P^0(\bar{\boldsymbol{y}})$.

Proposition 5. *If* $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}})$ *provides a Pareto-optimal cut, then all dual values* $\bar{\boldsymbol{\nu}}^0$ *are minimal, i.e., they cannot be decreased without violating constraints (13b) or changing some other dual value* ν. *This implies that all components of* $\bar{\boldsymbol{\nu}}^0$ *can be assumed minimal.*

Proof. Proof We assume that $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}})$ provides a Pareto-optimal cut. We prove by contradiction. Assume that for one position $p_0 \in P^0(\bar{\boldsymbol{y}})$ the value $\bar{\nu}_{p_0}$ could be decreased to $\bar{\nu}_{p_0} - \varepsilon$ (for some $\varepsilon > 0$) so that another dual feasible solution $(\dot{\boldsymbol{\mu}}, \dot{\boldsymbol{\nu}})$ could be constructed. This solution has a $\boldsymbol{\nu}$-part that fulfill $\dot{\nu}_p = \bar{\nu}_p$ for all $p \in P \setminus \{p_0\}$ and $\dot{\nu}_{p_0} = \bar{\nu}_{p_0} - \varepsilon$. The requirement that constraint (12b) (equivalent to (13b)) still holds true for the new solution $(\dot{\boldsymbol{\mu}}, \dot{\boldsymbol{\nu}})$ implies an increase of the difference $\dot{\mu}_o - \dot{\mu}_d$ by the same value $\varepsilon = \varepsilon(1 - \bar{y}_{p_0})$.

Then, for all $\boldsymbol{y}^\star \in Y$ with $y^\star_{p_0} = 1$, i.e., $(1 - y^\star_{p_0}) = 0$, the relation

$$\begin{aligned}\bar{\mu}_o - \bar{\mu}_d + \sum_{p \in P} \bar{\nu}_p(1 - y^\star_p) &= \bar{\mu}_o - \bar{\mu}_d + \bar{\nu}_{p_0} \cdot 0 + \sum_{p \in P \setminus \{p_0\}} \dot{\nu}_p(1 - y^\star_p) \\ &< \dot{\mu}_o - \dot{\mu}_d + \dot{\nu}_{p_0} \cdot 0 + \sum_{p \in P \setminus \{p_0\}} \dot{\nu}_p(1 - y^\star_p) = \dot{\mu}_o - \dot{\mu}_d + \sum_{p \in P} \dot{\nu}_p(1 - y^*_p)\end{aligned}$$

holds. Since enforcing the collection from the position p_0 is never restrictive in the SPRP-SS, such a $\boldsymbol{y}^\star \in Y$ certainly exists. For all other $\boldsymbol{y} \in Y$ with $y_{p_0} = 0$, i.e., $(1 - y_{p_0}) = 1$, we get

$$\begin{aligned}\bar{\mu}_o - \bar{\mu}_d + \sum_{p \in P} \bar{\nu}_p(1 - y_p) &= \bar{\mu}_o - \bar{\mu}_d + \bar{\nu}_{p_0} + \sum_{p \in P \setminus \{p_0\}} \dot{\nu}_p(1 - y_p) \\ &= \dot{\mu}_o - \dot{\mu}_d + \dot{\nu}_{p_0} + \sum_{p \in P \setminus \{p_0\}} \dot{\nu}_p(1 - y_p) = \dot{\mu}_o - \dot{\mu}_d + \sum_{p \in P} \dot{\nu}_p(1 - y^*_p)\end{aligned}$$

contradicting with the assumption that $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}})$ provides a Pareto-optimal cut.

Remark: For a feasible dual solution $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}})$ in which at only one position $p_0 \in P^0(\bar{\boldsymbol{y}})$ the value $\bar{\nu}_{p_0}$ can be decreased, we are able to interpret the function of the MW constraint in the SPRP-SS: Without loss of generality, we can assume that $\bar{\boldsymbol{\mu}}$ has been obtained from $\bar{\boldsymbol{\nu}}$ via Proposition 4. The decrease to the value $\bar{\nu}_{p_0} - \varepsilon$ increases the length of all edges that are in conflict with p_0 by ε. Since all the longer edges are at the same stage, the length of any path in $(V, E, \tilde{c})$ can only increase by not more than ε. The MW constraint imposes that the length $\bar{\mu}_o - \bar{\mu}_d$ of the optimal o-d-path increases by exactly ε. Hence, a decrease is only possible for positions p_0 covered by all shortest paths, and the decrease of $\bar{\nu}_{p_0}$ is limited to the point when there exists an optimal o-d-path which does not contain any of the effected edges. □

With all these results, we finally re-write the MW objective (12a) and (13a) into:

$$z_{\text{MW}(\bar{\boldsymbol{y}}, \tilde{\boldsymbol{y}})} = \max \underbrace{\mu_o - \mu_d + \sum_{p \in P} \nu_p}_{=\text{const}(\boldsymbol{\mu}, \boldsymbol{\nu})} \quad - \underbrace{\sum_{p \in P} \nu_p \tilde{y}_p}_{\text{weighted sum of } \boldsymbol{\nu}} \tag{14}$$

Observation 2. *The MW subproblem (12) pushes the dual prices of $\boldsymbol{\nu}^0 = (\nu_p)_{p \in P^0(\bar{\boldsymbol{y}})}$ towards $-\infty$.*

Moreover, for a dual optimal solution $(\boldsymbol{\mu}, \boldsymbol{\nu})$ for which the constant part const$(\boldsymbol{\mu}, \boldsymbol{\nu})$ of the optimality cut is maximized and kept fixed, the MW objective (12) pushes a weighted sum of the dual prices of $\boldsymbol{\nu} = (\nu_p)_{p \in P}$ towards $-\infty$. Though it looks like we were decreasing the dual prices of $\boldsymbol{\nu}^0$ and $\boldsymbol{\nu}^1$. However, dual prices $\boldsymbol{\nu}^1$ cannot be decreased without changing the constant part of the optimality cut. This is because, since we want to preserve the constant, any decrease in ν_p with $p \in P$ has to increase the difference $\mu_o - \mu_d$ by the same value. For variables of type $\boldsymbol{\nu}^1$ this is not possible, since we assumed dual optimality. Summarizing, the MW subproblem tends to provide larger coefficients for variables y_p with $p \in P^0(\bar{\boldsymbol{y}})$.

Example 3. *(continued from Examples 1 and 2) We now consider the master solution $(\bar{y}_0, \bar{y}_7) = (1, 0)$ (the blue tour in Figure 1a). By Proposition 1, the dual optimal solution with $\bar{\nu}_0 = -R$ and $\bar{\nu}_7 = 0$ results in $\bar{\mu}_d = -34$ yielding the optimality cut $z \geq -16 + 50y_0$. Using Observation 1, we set $\bar{\nu}_0 = 0$, since any value below 0 does not improve the objective. As a result, we obtain an improved optimality cut $z \geq 34$.*

Observation 2 allows us now to push the value of ν_7 below 0, as long as an equivalent decrease in $\bar{\mu}_d$ can be realized. The resulting values are $(\bar{\nu}_0, \bar{\nu}_7) = (0, -2)$ and $\bar{\mu}_d = -36$. This further improves the optimality cut resulting in $z \geq 34 + 2y_7$, strictly dominating the two aformentioned optimality cuts.

Propositions 3 and 5 can be used to increase the quality of optimality cuts (10a) in the SPRP-SS. Improvements can be established by moving single dual values while keeping the others fixed. The result is however sequence-dependent, i.e., the sequence in which dual variables are to be moved is unclear. We also did not find an efficient and practicable procedure to set the $\boldsymbol{\mu}$-variables and $\boldsymbol{\nu}$-variables simultaneously. Already for the $\boldsymbol{\mu}$-variables, it is not clear how these must be set (using forward and backward labeling gives bounds for the $\boldsymbol{\mu}$-values; on the shortest path, values are fixed). Even worse, the choice of these values influences the possible choices for the $\boldsymbol{\nu}$-values. We think that there is no simple combinatorial approach to solve the MW subproblem in the SPRP-SS case. These observations motivate the further procedure to generate high quality cuts.

4. Core-Maximal Cuts

We now transfer this idea to enhance optimality cuts for the SPRP-SS to the general case. With Observations 1 and 2 in mind, we propose to

(1) maximize the constant part $\text{const}(\boldsymbol{\pi}) = \boldsymbol{\pi}^\top \boldsymbol{b}$ of optimality cut (5), and

(2) maximize the coefficients of the variable part $-\boldsymbol{\pi}^\top F$ of optimality cut (5).

In which context is it is beneficial to generate optimality cuts that have a maximal constant? We would like the constant to provide a lower bound to each and every master solution $\boldsymbol{y} \in Y$. In the PB formulation of the SPRP-SS, i.e., model (6), this is the case because all $\boldsymbol{y}$-variables have a non-negative coefficient in the optimality cut (10a), which results from the negative sign of the variables in the corresponding $\leq$-constraints (6d) imposing that all dual $\boldsymbol{\nu}$-variables are non-positive. Examples 1–4 show this. A transformation into the desired form is straightforward for many problems. We show this for the CFLP and FCNFP in Appendix D and Appendix E. While the generation of core-maximal cuts is generally applicable, we expect them to be most influential in the described cases.

We can solve the following three LPs in sequence in order to yield (1) and (2):

Dual subproblem:

$$z_{\text{SP}(\bar{\boldsymbol{y}})} = \max \boldsymbol{\pi}^\top (\boldsymbol{b} - F\bar{\boldsymbol{y}})$$
s.t. (4b) and (4c)

Auxiliary dual subproblem 1:

$$z_{\text{Aux1}(\bar{\boldsymbol{y}})} = \max \boldsymbol{\pi}^\top \boldsymbol{b} \quad (15a)$$
$$\text{s.t. } \boldsymbol{\pi}^\top (\boldsymbol{b} - F\bar{\boldsymbol{y}}) = z_{\text{SP}(\bar{\boldsymbol{y}})} \quad (15b)$$
$$(4b) \text{ and } (4c) \quad (15c)$$

Auxiliary dual subproblem 2:

$$\max -\boldsymbol{\pi}^\top F \mathbf{1} \quad (16a)$$
$$\text{s.t. } \boldsymbol{\pi}^\top (\boldsymbol{b} - F\bar{\boldsymbol{y}}) = z_{\text{SP}(\bar{\boldsymbol{y}})} \quad (16b)$$
$$\boldsymbol{\pi}^\top \boldsymbol{b} = z_{\text{Aux1}(\bar{\boldsymbol{y}})} \quad (16c)$$
$$(4b) \text{ and } (4c) \quad (16d)$$

The first LP is the above dual subproblem DSP($\bar{\boldsymbol{y}}$) stated in (4) with $z_{\text{SP}(\bar{\boldsymbol{y}})} = z_{\text{DSP}(\bar{\boldsymbol{y}})}$ and (4b) and (4c) ensuring $A^\top \boldsymbol{\pi} \leq \boldsymbol{c}, \boldsymbol{\pi} \geq \mathbf{0}$, i.e., dual feasiblty. As suggested by Magnanti and Wong (1981), we guarantee dual optimality in the two auxiliary subproblems with a constraint of the form $\boldsymbol{\pi}^\top (\boldsymbol{b} - F\bar{\boldsymbol{y}}) = z_{\text{SP}(\bar{\boldsymbol{y}})}$, i.e., (15b) and (16b). The first auxiliary subproblem (15) is derived from Observation 1: Note that the term $\boldsymbol{\pi}^\top \boldsymbol{b}$ in the objective (15a) is the constant part of the associated optimality cut (5).

The second auxiliary subproblem (16) is derived from Observation 2: Note first that the variable part of the optimality cut is $\text{var}(\boldsymbol{\pi}) = -\bar{\boldsymbol{\pi}}^\top F \boldsymbol{y}$. For a given dual optimal solution and a given maximal constant part, i.e., (16b) and (16c), this auxiliary problem *max*imizes the sum of the coefficients of the $\boldsymbol{y}$-variables in the derived optimality cut (5). This is equivalent to *min*imzing the value of the dual variables, while keeping the objective value of the subproblem and the constant part of the optimality cut fixed. For the SPRP-SS, only the dual variables of type $\boldsymbol{\nu}^0$ fall into this category, see Section 3.4.

Definition 1. *A* core-maximal cut *(CMC) is an optimality cut that is derived from a dual solution to the dual subproblem* (16)*, which in turn results from the solution of the two subproblems problems* (4) *and* (15)*.*

4.1. Properties of Core-Maximal Cuts

For what follows, we denote the constant part of the associated optimality cut (5) as $\text{const}(\bar{\boldsymbol{\pi}}) = \bar{\boldsymbol{\pi}}^\top \boldsymbol{b}$. We highlight the defining property of core-maximal cuts and the difference to MW optimality cuts.

Proposition 6. *Let $\bar{\boldsymbol{y}} \in Y$ be a solution to the Benders master problem. Let $\bar{\boldsymbol{\pi}}$ be a dual optimal solution from which the core-maximal cut* (5) *is derived. Then,*

(i) the constant const($\bar{\boldsymbol{\pi}}$) is maximal among all dual optimal solutions;

(ii) fixing the maximal constant, the sum of all $\boldsymbol{y}$-coefficients is maximized;

(iii) the cut derived from $\bar{\boldsymbol{\pi}}$ is not necessarily Pareto-optimal, i.e., not necessarily a Magnanti and Wong cut;

(iv) a Pareto-optimal cut derived from $\boldsymbol{\pi}^\star$ does not necessarily maximize its constant part const($\boldsymbol{\pi}^\star$), i.e., the cut is not necessarily a core-maximal cut.

Proof. Proof (i) and (ii): Direct by construction via solving models (4), (15), and (16).
(iii) and (iv): see Example 4 below or Appendix B for more general ideas. □

Example 4. *(continued from Example 3) Example 3 explains the different behavior of CMC and MW optimality cuts. Recall that the dual values $\bar{\mu}_o = 0, \bar{\mu}_d = -34, \bar{\nu}_0 = 0$, and $\bar{\nu}_7 = -2$ yields the core-maximal cut $z \geq 34 + 2y_7$. Consider the inner point $\tilde{\boldsymbol{y}}$ with $\tilde{y}_0 = \tilde{y}_7 = 0.8$, and $\tilde{y}_p = 1$ for $p \in P^{req}$ (otherwise). The MW subproblem then yields the optimal values $\bar{\mu}_o = -0, \bar{\mu}_d = -48, \bar{\nu}_0 = -8$, and $\bar{\nu}_7 = -10$ resulting in the MW optimality cut $z \geq 26 + 8y_0 + 10y_7$. This cut does not provide the largest constant among all dual optimal solutions. Therefore, the MW cut $z \geq 26 + 8y_0 + 10y_7$ is not a core-maximal cut.*

However, the MW cut dominates the core-maximal cut in the Pareto sense. To see this, we consider all three master solutions $\boldsymbol{y} \in Y$ with $(y_0, y_7) = (0,1)$, $(1,0)$, and $(1,1)$. The core-maximal cuts evaluate to $z \geq 36$, $z \geq 34$, and $z \geq 36$, while the MW cuts evaluate to $z \geq 36$, $z \geq 34$, and $z \geq 44$, respectively. Hence, the MW optimality cut provides the same or, in one case, a tighter lower bound than the CMC. The example shows that the two subproblems aim at different optimization criteria, and can lead to different optimality cuts. Note that the shown example is rather specific, and we expect many core-maximal cuts to meet the Pareto-optimality criterion.

4.2. Boundedness of the Auxiliary Subproblem

A few remarks regarding the boundedness of the auxiliary subproblems (15) and (16) are due. First, we implicitly assumed, like Magnanti and Wong (1981) did, that the basic dual subproblem DSP($\bar{\boldsymbol{y}}$) is bounded with respect to the given BM solution $\bar{\boldsymbol{y}} \in Y$. Hence, the MW subproblem and our auxiliary subproblems (15) and (16) can only generate optimality cuts.

For Benders decomposition with MW cuts, we provide an example for an unbounded dual MW subproblem in Appendix C. The idea is that every point in the (relative) interior can be represented by the convex combination of extreme points of its polyhedron. Then, the MW objective function can be rewritten as a weighted sum of objectives corresponding to these extreme points of the master polyhedron. If all extreme points result in feasible primal subproblems, i.e. bounded dual subproblems, each weighted component of the MW objective function optimizes towards a bounded direction, and the MW subproblems is bounded. However, if the MW objective function has unbounded components with positive weight, it can be unbounded.

For Benders decomposition with CMCs, a similar problem can arise. Depending on the problem, the constant in the cut might be unbounded. In this case, our second LP (15) also becomes unbounded. In particular, this happens if the DSP has extreme rays that are neutral with respect to the objective but beneficial in terms of the constant.

4.3. A Single Equivalent Subproblem

We introduced models (4), (15), and (16), which constitutes a lexicographic multi-objective LP. Herein, the objective (4a) of the standard dual subproblem DSP($\bar{\boldsymbol{y}}$) has to be maximized first, then (15a) second, and (16a) third. While lexicographic multi-objective LPs can be solved *preemptively*, i.e., one LP after the other, the literature also proposes *non-preemptive* approaches with the new objective being a weighted sum of the old ones. Sherali and Soyster (1983) show that (in the linear case) for each *preemptive* problem with an optimal solution, there exists a set of weights for the *non-preemptive* problem, such that any optimal solution of the *non preemptive* problem is also optimal for the *preemptive* problem. Using the underlying ideas of Sherali and Soyster, we provide sufficient estimates for ε_1 and ε_2 in case of the SPRP-SS in Section 4.4 as well as for the CFLP in Appendix D and the FCNFP in Appendix E. The non-preemptive model is:

$$z_{\mathrm{CMC}(\bar{\boldsymbol{y}})} = \max \quad \boldsymbol{\pi}^\top(\boldsymbol{b} - F\bar{\boldsymbol{y}}) + \varepsilon_1 \cdot (\boldsymbol{\pi}^\top \boldsymbol{b}) + \varepsilon_2 \cdot (-\boldsymbol{\pi}^\top F\mathbf{1}) \tag{17a}$$

$$\text{subject to} \quad A^\top \boldsymbol{\pi} \le \boldsymbol{c}, \boldsymbol{\pi} \ge 0 \tag{17b}$$

If the values $0 < \varepsilon_2 \ll \varepsilon_1 \ll 1$ are chosen appropriately, the model (17) solves the dual subproblem and produces core-maximal cuts. In general, the values of ε_1 and ε_2 must be chosen sufficiently small so that the maximum contribution of the term scaled by an ε-value is always smaller than the minimum contribution of the prior term. In particular, the maximum gain in the last term must be smaller than ε_1, which is, in turn, less than the minimum loss in the first term. Specific ε-values must be computed for each type of problem and the instance of the problem at hand.

4.4. Choice of the ε-Values in the SPRP-SS

For the SPRP-SS, the objective (17a) can be re-written so that the coefficients of the dual variables $(\boldsymbol{\mu}, (\boldsymbol{\nu}^0, \boldsymbol{\nu}^1))$ become clear. The CMC subproblem then reads as follows:

$$z_{\mathrm{CMC}(\bar{\boldsymbol{y}})} = \max \quad (1+\varepsilon_1)(\mu_o - \mu_d) \quad + \quad (1+\varepsilon_1-\varepsilon_2)\cdot \sum_{p\in P^0(\bar{\boldsymbol{y}})} \nu_p \quad + \quad (\varepsilon_1-\varepsilon_2)\cdot \sum_{p\in P^1(\bar{\boldsymbol{y}})} \nu_p \tag{18a}$$

$$= \max \quad (1+\varepsilon_1)\Big(\mu_o - \mu_d + \sum_{p\in P^0(\bar{\boldsymbol{y}})} \nu_p\Big) \quad - \quad \varepsilon_2 \cdot \sum_{p\in P^0(\bar{\boldsymbol{y}})} \nu_p \quad + \quad (\varepsilon_1-\varepsilon_2)\cdot \sum_{p\in P^1(\bar{\boldsymbol{y}})} \nu_p \tag{18b}$$

$$\text{subject to} \quad (9b) \text{ and } (9c) \tag{18c}$$

We can interpret the objective (18a) as follows: For the $\boldsymbol{\nu}^1$-variables, Observation 1 is enforced due to coefficients $(\varepsilon_1 - \varepsilon_2)$ that are slightly greater than zero. Since they now contribute to the objective, they become as large as possible, i.e., they are pushed towards zero. The coefficients $(1+\varepsilon_1-\varepsilon_2) > 1$ of the $\boldsymbol{\nu}^0$-variables are slightly smaller than the coefficients $(1+\varepsilon_1) > 1$ of the $\boldsymbol{\mu}$-variables. It is therefore beneficial to first increase the difference $\mu_o - \mu_d$.

Theorem 1. *If the weights $\varepsilon_1 > \varepsilon_2 > 0$ fulfill $\varepsilon_1 < 1/(|P|\cdot R)$ and $\varepsilon_2 < \varepsilon_1/(|P|\cdot R+1)$, then model (17) solves the lexicographic multi-objective LP given by (4), (15), and (16) for the SPRP-SS.*

Proof. Proof For a formal derivation, we will use the following facts:

(a) Solutions to the Benders subproblems have integral costs $z_{\mathrm{SP}(\bar{\boldsymbol{y}})} = z_{\mathrm{DSP}(\bar{\boldsymbol{y}})} \in \mathbb{Z}_+$ because of the integer cost-coefficient assumption and Proposition 1.

(b) The total variation in $\boldsymbol{\nu}$-variables is upper bounded by $|P|\cdot R$, as was also shown in Proposition 1(c).

The proof has three parts corresponding to the three LPs that have to be solved in a lexicographic fashion:

1. We want to show that model (18) provides a dual optimal solution, i.e., an optimal solution to (9). The maximum variation in the second and third sum of (18b) is upper bounded by $\varepsilon_1|P|\cdot R$ due to fact (b). For $\varepsilon_1|P|\cdot R < 1$ which is equivalent to $\varepsilon_1 < 1/(|P|\cdot R)$, the maximum total variation in the second and third sum in (18b) is less than the minimum change in $z_{\mathrm{SP}(\bar{\boldsymbol{y}})}$ due to fact (a). Therefore, (18) maximizes the first term, i.e., maximizes $\mu_o - \mu_d + \sum_{p\in P^0(\bar{\boldsymbol{y}})} \nu_p$, which is equal to $z_{\mathrm{SP}(\bar{\boldsymbol{y}})}$ in the maximum.
2. We want to show that (18) maximizes the constant part of the optimality cut while preserving dual optimality. In 1. we have shown that the first term is maximal so that $(1+\varepsilon_1)z_{SP(\bar{\boldsymbol{y}})}$ is constant and fixed. Next, we compare the second and the third term: The maximum total variation in the second term is upper bounded by $\varepsilon_2|P|\cdot R$. For $\varepsilon_2|P|\cdot R < \varepsilon_1 - \varepsilon_2$ which is equivalent to $\varepsilon_2 < \varepsilon_1/(|P|\cdot R+1)$, it follows that model (18) maximizes the third term, i.e., maximizes $\sum_{p\in P^1(\bar{\boldsymbol{y}})} \nu_p$. Since $\mathrm{const}(\boldsymbol{\mu},\boldsymbol{\nu}) = z_{\mathrm{SP}(\bar{\boldsymbol{y}})} + \sum_{p\in P^1(\bar{\boldsymbol{y}})} \nu_p$ for a dual optimal solution, the result is proven.
3. We want to show that model (18) maximizes the sum of all coefficients of $\boldsymbol{y}$ in the optimality cut, while preserving dual optimality and maximality of the constant part of the cut. Recall from (10b) that $\mathrm{var}(\bar{\boldsymbol{\nu}}) := -\sum_{p\in P} \bar{\nu}_p y_p$. In 1. and 2. we have shown that model (18) maximizes $z_{\mathrm{SP}(\bar{\boldsymbol{y}})}$ and $\mathrm{const}(\boldsymbol{\mu},\boldsymbol{\nu})$, respectively. Hence the first term and the third term in (18b) are fixed. Therefore, (18b) maximizes the remaining coefficient of $\mathrm{var}(\bar{\boldsymbol{\nu}})$ that are not already fixed due to dual optimality and maximality of the constant part. □

Note that the estimates used in the proof of Theorem 1 are very generous and in fact, tighter bounds for the maximum variation can be derived. Therefore, in our computational study, larger values were sufficient, e.g., $\varepsilon_1 = 10^{-3}$ and $\varepsilon_2 = 10^{-5}$.

5. Computational Results

In this section, we compare various implementations of Benders decomposition algorithms with MIP-based solution approaches that solve a given formulation directly. For the former, we compare different types of Benders optimality cuts, including the new core-maximal cuts. We present detailed results for the SPRP-SS, and we also summarize results for two other prominent problems for which Benders algorithms have been applied successfully in the literature: the CFLP and the FCNFP. Section 5.1 describes the SPRP-SS benchmark instances, Section 5.2 describes the different approaches, and Section 5.3 describes the computational setup. Subsequently, we present and discuss the actual comparison based on runtimes and the number of Benders iterations: for the SPRP-SS in Section 5.4, for the CFLP in Section 5.5, and for the FCNFP in Section 5.6.

5.1. SPRP-SS Benchmark Instances

We generate benchmark instances as described in (Goeke and Schneider, 2021; Heßler and Irnich, 2024; Lüke et al., 2024). These instances are characterized by a combination of four parameters (m, C, a, α), where m denotes the number of aisles, C denotes the number of cells per aisle, a denotes the number of different articles to collect, and α denotes the scattering policy. We consider single-block, rectangular warehouses with $m = 10$ aisles and $C = 50$ cells per aisle (note that every two opposite cells per aisle result in the same pick position). We will consider instances with pick list of size $a \in \{30, 50, 100, 200\}$. Since warehouses differ in storage and scattering strategies, the experiments cover a broad range of possibilities. The parameter α describes one of the following five settings:

- To account for cases where the scattering is completely independent of the article type, we use two uniform scatter factors $\alpha = 1.5$ and $\alpha = 2.5$. This indicates that, on average, each article can be found in the warehouse at a random position 1.5 times or 2.5 times, respectively.
- Articles can often be categorized by sales volume using an ABC classification, where 20% are category A (highest sales volume), 30% are category B, and 50% are category C (lowest sales volume). Accordingly, the probability of an article occurring in a typical pick list depends on its category, with 80%, 15%, and 5% probability, respectively. Lüke et al. (2024) found that, for chaotic storage and an overall scatter factor of 2, scattering only category A articles with a factor of 6, while not scattering articles from the other two categories, leads to the shortest picker tours on average. We denote this setting by $\alpha = 6/1/1$.
- The last two cases cover warehouses in which a significant portion of the articles are not scattered at all, though some are stored in many different locations. Therefore, we analyze two settings. In the first one, only 25% of the articles are scattered (75% have unique storage positions), but they are scattered at a high rate of 30, resulting in a total scatter factor of 8.25. We denote this setting by $\alpha = 25\%/30$. In the second setting, 50% of the articles are scattered with a high factor of 15, resulting in a total scatter factor of 8. We denote this setting by $\alpha = 50\%/15$.

The scattering of the SKUs is implemented as follows. First, for each article $s \in S$, we uniformly randomly draw a pick positions $p \in P$. This defines a pair $(s, p) \in S \times P$. After this procedure, every article is available exactly once in the warehouse. Second, $(\alpha - 1) \cdot a$ unique pairs $(s, p) \in S \times P$ are uniformly randomly drawn (we repeat drawing if a pairs is generated more than once). The latter guarantees that on average articles are available exactly α times. In case of two or three categories (ABC classification or scattered/not scattered), the second step is performed independently for each category of articles. All these generated pairs (s, p) together define the sets P_s of potential pick positions for all articles $s \in S$.

In the unit-demand case, demand and supply are set to $q_s = 1$ for all $s \in S$ and $b_{sp} = 1$ for all $p \in P_s$. In general-demand case, each pair (s, p) is assigned to a random supply $b_{sp} \in \{1, 2, 3\}$. Demands q_s range from $\{1, 2, ..., \min\{6, \sum_p b_{sp}\}\}$ making all instances feasible by construction.

The benchmark set that we use consists of 50 instances per combination (m, C, a, α), yielding a total benchmark set of $2 \cdot 4 \cdot 5 \cdot 50 = 2000$ instances. The instances are available at `https://logistik.bwl.uni-mainz.de/research/benchmarks/`.

5.2. Compared Approaches

The state-of-the-art approach for the SPRP-SS is to use an IP or MIP formulation tailored to the problem, which is then solved directly with an MIP solver. This approach is also standard for the CFLP and FCNFP. We compare:

- HI: the network-flow formulation of Heßler and Irnich (2024)
- RG: the improved formulation of Lüke et al. (2025), particularly suited for SPRP-SS instances with massive scattering
- PB: the new positional-based formulation (6) presented in Section 3.2
- MIP: two standard MIP formulations for the CFLP and the FCNFP, see Appendix D and Appendix E

For implementing Benders decomposition algorithms, there exist various options. We compare:

- CPX-Benders: automatic Benders decomposition available in `CPLEX`
- DSP-direct: direct solution of the *dual subproblem* (DSP) given by (9) using dynamic programming labeling and Proposition 1. Since the choice of $M = -R$ leads to weak cuts (compare Proposition 3), we compute position-specific values of ν_p for $p \in P$ in a greedy fashion to set them as large as possible (close to 0) while not violating constraints.
- DSP-warm: solution of DSP using the LP solver of `CPLEX`, warm-started with the previous LP solution from the preceding iteration (this is the default when a model is modified and solved again)
- DSP-reset: solution of DSP using the LP solver of `CPLEX`, but resetting dual variables to 0 (see below)
- MW: MW cuts computed with subproblem (12) presented in Section 3; the value $z_{\mathrm{DSP}(\bar{\boldsymbol{y}})}$ is calculated with a labeling algorithm using Proposition 1
- MND: *maximal nondominated cuts* (MNDs) computed with the subproblem of Sherali and Lunday (2013)
- CMC: core-maximal cuts computed with the subproblem (18) presented in Section 4.4

Regarding the warm-start of the LP solver of `CPLEX`, we see that more dual variables ν_p for $p \in P$ assume non-zero values as when solving every DSP from scratch. Therefore, the result in one Benders iteration is trajectory-dependent in algorithm DSP-warm, and its behavior is somewhat uncontrolled. To eliminate this effect, we also implemented the DSP optimization in algorithm DSP-reset, which starts with a penalty on the $\boldsymbol{\nu}^0$-variables, resetting them to zero when possible.

We also tested several other methods for generating the cuts that are presented in literature. Stursberg (2019) introduces facet-defining cuts which are based on the MIS cuts from Fischetti et al. (2010). For the SPRP-SS, pretests revealed that neither type of optimality cut is competitive with the ones listed above. Therefore, they are not included in the computational study. Since our subproblems differ very much from the ones where OLS and MIS are benefitial, we did not implement a Benders algorithm using OLS cuts. Saharidis et al. (2010) formalized the idea of generating multiple cuts per iteration. The proposed covering cut bundle method generates several cuts per iteration. However, pre-tests also showed that this technique is not beneficial for the SPRP-SS. This is because the bottleneck of our Benders procedure is solving the Benders master problem iteratively, which becomes increasingly more difficult with the number of optimality cuts included. Thus, any method that generates multiple cuts per iteration was outperformed by adding only a single cut per iteration (we also tested different combinations of different cuts, e.g., two MW cuts etc.). A straightforward implementation of the idea of adding cuts before branch-and-cut, as proposed by Fischetti et al. (2016), did not improve the performance in our case.

5.3. Computational Setup

All algorithms are implemented in `C++` using the concert API of `CPLEX` 22.1.0 and are compiled into 64-bit single-thread release code with Microsoft Visual Studio 2022. Default parameters of `CPLEX` are kept, except for setting the number of available threads to one. Moreover, the maximum computation time per instance is limited to 600 seconds (10 minutes). The computational study is performed on a 64-bit Microsoft Windows 10 computer equipped with an Intel® Core™ i7-5930k CPU clocked at 3.5GHz and 64GB of RAM.

5.4. Comparison of Computational Performance for the SPRP-SS

Table 1 shows average computation times (in milliseconds) grouped by the five scattering scenario, the demand scenario (unit demand and general demand), and the number a of different articles to collect. The fastest runtime in each of the 40 groups is highlighted in bold. The automatic Benders decomposition, algorithm CPX-Benders, is an out-of-the-box Benders implementation provided by `CPLEX` based on our PB formulation. We do not know details of this solution procedure, in particular, it is not clear what type of Benders optimality cuts and cut separation technique are used. We observed however that `CPLEX` sometimes switched back to regular MIP solving using traditional cuts and branch-and-bound. For a few instances of the 25%/30 scattering settings, the algorithm CPX-Benders crashed (with the message that presolving results in an empty Benders master; for all 50 instances with unit demand and $a = 200$). We could suppress these crashes by turning presolving off, which however would increase runtimes substantially. Instead of using this option, we have simply omitted these results when computing the runtime averages, and we indicate these events by attaching an asterisk * to the respective value.

Comparing the three MIPs with the seven Benders algorithms, the average runtime of the MIP solver is consistently below one second, hovering around a few hundred milliseconds. As expected, formulation RG outperforms formulation HI when a larger number a of articles must be collected, and a similar behavior can be observed comparing formulation PB with both formulations HI and RG. In contrast, the runtime of the Benders algorithms varies greatly, ranging from less than ten milliseconds and the time limit of 600 seconds.

Table 1: Average runtime (in milliseconds) of the compared approaches for the SPRP-SS. The fastest runtime in each group is indicated with bold numbers. An asterisk indicates an internal error (for 1, 30, 5, 11, 15, and 50 of the 50 instances, respectively) in `CPLEX`'s automatic Benders algorithm.

		general demand				unit demand			
		number of different articles				number of different articles			
scattering setting	algorithm	$a = 30$	50	100	200	$a = 30$	50	100	200
$\alpha = 1.5$	HI	11.8	15.2	21.2	32.7	12.7	16.6	26.8	40.1
	RG	12.7	15.1	19.7	27.2	13.3	16.3	27.3	40.7
	PB	7.2	8.6	11.6	14.8	10.0	13.0	19.9	30.0
	CPX-Benders	12.2	14.5	19.6	24.8	20.0	27.8	44.5	61.5
	DSP-direct	7.3	14.7	14.4	12.7	130.9	1628.5	70007.9	32162.5
	DSP-warm	3.8	3.9	4.7	5.4	7.5	**6.8**	7.9	8.6
	DSP-reset	18.5	33.6	32.9	33.2	116.6	546.5	18969.3	4252.1
	MW	3.9	4.4	5.5	8.2	10.9	11.1	10.1	10.9
	MND	3.5	3.9	5.1	8.0	7.3	7.5	8.0	10.6
	CMC	**3.3**	**3.4**	**3.9**	**4.4**	**6.4**	**6.8**	**6.7**	**6.7**
$\alpha = 2.5$	HI	24.3	34.1	60.3	123.4	37.3	60.5	142.5	369.5
	RG	30.7	41.8	70.9	135.9	**35.7**	**48.6**	**95.0**	181.3
	PB	15.5	20.8	31.8	47.4	38.4	59.0	108.4	165.1
	CPX-Benders	33.6	46.8	68.2	98.2	142.3	264.3	582.6	1113.7
	DSP-direct	1862.7	76963.5	151681.9	37549.4	497423.2	579155.5	600000.0	600000.0
	DSP-warm	10.5	10.8	9.7	9.8	185.2	467.4	142.1	99.9
	DSP-reset	424.9	15909.7	45233.4	3645.4	277018.2	483619.9	576739.4	590873.1
	MW	19.7	16.0	12.1	18.2	553.0	1093.9	556.2	227.8
	MND	11.1	10.6	11.0	19.0	181.2	325.4	121.4	95.1
	CMC	**10.1**	**8.5**	**7.5**	**9.1**	133.8	299.6	110.6	**70.2**
$\alpha = 6/1/1$	HI	145.5	266.7	609.3	972.5	66.4	115.2	259.7	351.7
	RG	162.1	294.6	288.9	469.4	**53.3**	**78.3**	**129.0**	261.7
	PB	**115.4**	**178.0**	221.7	168.8	114.3	199.4	304.8	273.1
	CPX-Benders	494.3	1340.3	1689.4	1009.2	1036.8	1220.1	2618.4	1569.6
	DSP-direct	574532.7	600000.0	600000.0	600000.0	600000.0	600000.0	588089.3	318176.3
	DSP-warm	3359.6	12585.1	171.2	60.9	12285.6	1162.0	166.6	68.8
	DSP-reset	516477.9	599443.9	593252.3	595504.7	525757.6	573710.8	555484.6	197089.4
	MW	5258.5	12382.7	557.6	105.1	15317.7	3590.0	846.5	167.3
	MND	1531.3	6998.3	161.1	73.6	7866.2	949.1	207.3	99.6
	CMC	1117.2	2315.1	**137.6**	**41.9**	4976.3	364.3	172.6	**69.2**
$\alpha = 25\%/30$	HI	41.7	67.6	123.5	173.6	**23.9**	45.5	95.6	127.5
	RG	111.7	189.2	313.2	461.4	31.7	51.1	111.8	308.5
	PB	110.9	155.5	170.3	94.1	109.2	153.8	150.5	91.3
	CPX-Benders	531.8	635.5	*353.3	*65.6	*307.8	*289.7	*88.0	*—
	DSP-direct	36936.9	59.9	37.8	**7.3**	37.7	**25.4**	**16.0**	**6.0**
	DSP-warm	30.2	36.8	36.2	22.9	25.7	31.2	30.3	22.0
	DSP-reset	748.2	320.7	241.0	61.9	172.4	154.6	96.7	45.2
	MW	96.7	72.7	41.0	35.0	53.1	50.9	43.7	41.1
	MND	35.4	39.0	34.7	31.8	29.1	34.0	34.4	35.9
	CMC	**33.3**	**33.6**	**27.7**	19.8	25.3	30.6	30.4	22.4
$\alpha = 50\%/15$	HI	94.7	194.4	556.9	1149.2	36.6	77.1	191.8	291.1
	RG	113.9	174.0	381.6	638.0	**36.0**	**57.1**	139.5	658.7
	PB	141.8	208.8	290.1	214.2	133.4	213.1	303.6	220.9
	CPX-Benders	847.0	1658.3	1843.2	758.2	623.9	1067.4	1059.1	395.3
	DSP-direct	444079.4	460201.4	92995.8	137.1	155520.0	48980.8	86.6	**32.4**
	DSP-warm	73.4	74.1	65.5	39.3	40.1	54.3	59.9	38.1
	DSP-reset	278162.5	260378.7	46319.8	1841.2	29542.9	7848.6	864.7	495.0
	MW	368.6	289.7	185.5	64.1	146.5	169.4	122.6	58.5
	MND	85.0	91.1	95.2	71.7	48.1	67.5	78.2	56.4
	CMC	**71.9**	**79.2**	**71.5**	**35.4**	42.2	60.4	**60.9**	37.3

Clearly, Benders algorithms' performance depends heavily on the cut selection strategy. The overall trend is that Algorithm MND slightly outperforms algorithm MW, in particular for the last three scattering settings $\alpha = 6/1/1$, 25%/30, and 50%/15, while the algorithm CPX-Benders with automated Benders decomposition is not competitive in comparison to the two former algorithms. The Benders algorithm CMC outperforms all three above algorithms. It achieves the fastest runtime in 24 out of 40 groups.

We discuss the results for the different scattering settings now:

- For $\alpha = 1.5$ (where on average only every second article of the pick list has two possible collection positions), the group of instances is clearly the most easy for all types of algorithms, resulting in the smallest runtimes in comparison. This can be explained by the fact that in these instances, a substantial part of the tour is predetermined by the articles that have a unique pick position. This predetermined part may already collect most of the scattered articles. Compared to the other scattering settings, the MIP formulations tend to have a smaller number of variables, and the Benders algorithms have optimality cut that tend to provide more information useful for subsequent master solutions. Only algorithms DSP-direct and DSP-reset have average runtimes above one second, which happens in case of unit demand and for $a \geq 50$.
- For $\alpha = 2.5$, this group of instances is more difficult than for $\alpha = 1.5$, resulting in systematically longer runtimes. For general demand, the Benders algorithms perform very well (except for DSP-direct and DSP-reset), outperforming all direct MIP approaches. For unit demand, this is the only group for which MIP-based algorithms (HI, RG, and PB) are significantly faster than all Benders algorithms (except for $a = 200$ where algorithms MND and CMC are faster). Comparing general and unit demand, the runtime of Benders algorithms increases by a factor of at least ten, while the algorithms HI, RG, and PB remain consistent in runtime. Comparing different numbers a of articles in the pick list, MIPs tend to become slower due to an increasing number of variables and constraints, while Benders algorithm tend to become faster. Note that with increasing a also the number of articles being available at only one position increases, which makes instances easier for Benders algorithms.
- For $\alpha = 6/1/1$, we can see that this setting is the most challenging for Benders algorithms. Especially for small $a = 30$ and 50, Benders algorithms are over 20 times slower than the MIPs. As before, DSP-direct and DSP-reset are not at all competitive, often reaching the runtime limit of 600 seconds for (almost) all instances in a group. While the other Benders algorithms catch up with the MIPs in the group for $a = 200$, our Benders algorithm with the new core-maximal cuts is more than competitive already for $a = 100$, and it is at the same time substantially faster than the other Benders algorithms. We can explain the general behaviour of the algorithms: Recall that in this scattering setting articles of category A make up 80% of the pick list, each being scattered six times on average. As a result, the majority of the articles is scattered multiple times so that the Benders master has a huge number of different, good solutions. With increasing a however, the pick list tends to comprise more articles with a unique pick position, making instances easier to solve for Benders algorithms.

  We also analyzed the variation of runtimes within the eight groups for $\alpha = 6/1/1$. The number of Benders iterations (see also analysis below) is highly instance-dependent, especially for smaller a-values. In comparison, Benders algorithms' runtimes are less stable compared to the direct MIP-based approaches that have rather predictable runtimes. We saw, for many instances, that Benders algorithms have the smallest runtime, but the worst-case runtime is often long, distorting the average runtime presented in Table 1.

- For $\alpha = 25\%/30$, we can note that 75% of the articles to collect have a unique pick position, resulting in large parts of the tour being predetermined. Again, this property makes the instances relatively easy for all types of solution approaches. That CPLEX's automatic Benders algorithm crashes for some instances is probably caused by this fact. Interestingly, the direct solution of the DSP without LP solver in algorithm DSP-direct is a competitive approach here. In four of the eight groups, algorithm DSP-direct is faster than all other algorithms (while it performs worst for general demand and $a = 30$). In three of the remaining cases, however, our new algorithm CMC is the fastest, showing a consistent average runtime below 35 milliseconds in all eight groups.
- For $\alpha = 50\%/15$, a larger portion of the articles is scattered compared to setting $\alpha = 25\%/30$. Thus, instances become more difficult for Benders algorithms in comparison, while the performance of the MIP-based algorithms (HI, RG, and PB) deteriorates only slightly. The relative comparison between algorithms is similar to what we have seen in setting $\alpha = 25\%/30$.

For the 2000 benchmark instances, we provide detailed instance-specific results at https://logistik.uni-mainz.de. Overall, algorithm CMC is the fastest in 870 cases, algorithm DSP-warm in 368 cases, algorithm DSP-direct in 339 cases, algorithm RG in 214 cases, algorithm MND in 88 cases, algorithm HI in 56 cases, algorithm PB in 63 cases, and algorithm MW in 2 cases.

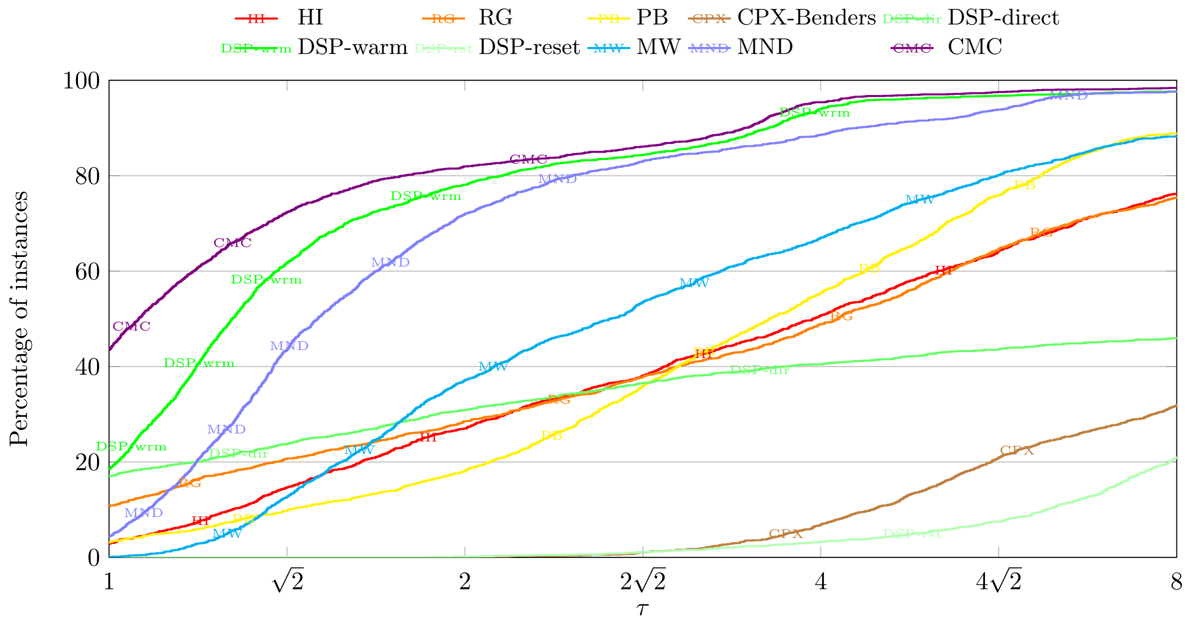


Figure 2: Performance profiles comparing the ten exact algorithms for the SPRP-SS.

The overall performance is visualized in the performance profile shown in Figure 2 (Dolan and Moré, 2002). Given the ten SPRP-SS algorithms $A_1, A_2, \ldots, A_{10}$ from Section 5.2, the performance profile $\rho_{A_k}(\tau)$ of algorithm A_k, $1 \leq k \leq 10$ describes the ratio of instances that can be solved by A_k within a factor τ compared to the fastest algorithm. Formally, the profile is defined by

$$\rho_{A_k}(\tau) = \left|\left\{ I \in \mathcal{I} : t_I^{A_k}/t_I^* \leq \tau \right\}\right| / |\mathcal{I}|,$$

where $\mathcal{I}$ is the benchmark set, $t_I^{A_k}$ is the computation time of algorithm A_k when applied to instance $I \in \mathcal{I}$, and t_I^* is the minimal computation time of all algorithms solving the instance I. Unsolved instances count with infinite time so that, for large values of τ, the value $\rho_{A_k}(\tau)$ is the percentage of instances solved by A_k within the time limit. On the other hand, $\rho_{A_k}(1)$ is the percentage of instances for which algorithm A_k is the fastest. The profiles confirm that algorithm CMC outperforms all other nine algorithms on the benchmark.

We also analyze the effectiveness of the different Benders optimality cuts by counting how many of them are necessary. Since all Benders algorithms add exactly one cut per iteration, the number of cuts is identical to the number of iterations. Table 2 shows the average number of Benders cuts generated until termination (possibly prematurely when the time limit of 600 seconds is reached).

Table 2: Average number of Benders optimality cuts of the compared approaches for the SPRP-SS. The smallest number of cuts is indicated in bold (disregarding algorithm CPX-Benders because of its uncontrollable behavior of falling back to the standard branch-and-cut MIP approach). An asterisk indicates an internal error in `CPLEX`'s automatic Benders algorithm.

		general demand				unit demand			
		number of different articles				number of different articles			
scattering setting	algorithm	$a = 30$	50	100	200	$a = 30$	50	100	200
$\alpha = 1.5$	CPX-Benders	1.7	1.9	1.8	1.7	3.9	4.6	3.8	1.0
	DSP-direct	15.1	27.6	19.9	10.7	130.8	480.7	3744.8	2154.5
	DSP-warm	3.8	3.9	4.1	3.8	6.5	5.9	5.7	5.2
	DSP-reset	11.5	17.7	11.8	8.7	62.4	157.2	1177.3	498.1
	MW	2.7	2.7	2.6	2.5	5.9	5.1	3.9	2.7
	MND	2.7	2.8	2.9	2.7	5.5	4.9	4.0	3.3
	CMC	**2.5**	**2.2**	**2.1**	**2.0**	**4.3**	**4.4**	**3.6**	**2.0**
$\alpha = 2.5$	CPX-Benders	6.5	7.4	4.0	4.0	31.6	36.7	27.3	24.0
	DSP-direct	549.0	3423.0	7609.0	2101.0	12897.0	12598.0	10856.0	10684.0
	DSP-warm	8.5	7.9	5.9	5.4	49.9	54.6	40.8	14.1
	DSP-reset	151.0	956.0	2764.0	457.0	4900.0	6804.0	8496.0	7914.0
	MW	8.0	5.9	3.4	2.9	70.6	82.6	42.8	14.1
	MND	6.6	5.6	3.9	3.8	49.8	52.6	29.7	13.6
	CMC	**6.3**	**4.4**	**2.5**	**2.0**	**42.0**	**41.3**	**28.1**	**12.1**
$\alpha = 6/1/1$	CPX-Benders	53.3	51.0	32.1	10.5	62.5	49.4	32.9	19.2
	DSP-direct	14516.0	13076.0	11814.0	14072.0	20734.0	16854.0	24483.0	8283.0
	DSP-warm	99.8	86.7	31.6	12.6	112.3	77.1	20.5	10.1
	DSP-reset	7814.0	8893.0	7740.0	6290.0	10385.0	9158.0	9084.0	1211.0
	MW	164.1	103.6	25.6	6.1	108.1	118.9	24.4	6.6
	MND	87.4	66.8	21.4	7.2	99.4	65.3	19.0	7.1
	CMC	**73.8**	**61.4**	**20.2**	**5.2**	**81.4**	**50.8**	**17.5**	**5.9**
$\alpha = 25\%/30$	CPX-Benders	16.5	12.0	*3.7	*1.0	*9.1	*6.4	*1.9	*—
	DSP-direct	1644.4	19.7	12.0	2.9	17.4	8.4	4.5	**2.0**
	DSP-warm	7.3	6.1	4.4	2.2	4.8	4.0	2.3	**2.0**
	DSP-reset	44.0	14.4	9.6	2.7	11.2	6.8	3.6	**2.0**
	MW	6.9	4.7	2.3	**2.0**	4.1	2.9	**2.1**	**2.0**
	MND	6.9	5.3	2.9	2.1	4.7	3.5	**2.1**	**2.0**
	CMC	**6.5**	**4.3**	**2.3**	**2.0**	**3.8**	**2.8**	**2.1**	**2.0**
$\alpha = 50\%/15$	CPX-Benders	27.0	23.8	16.9	2.5	20.8	17.4	10.9	1.5
	DSP-direct	10692.0	11448.0	2968.0	25.8	5063.0	2153.0	16.4	9.5
	DSP-warm	17.8	12.2	8.7	6.7	9.3	7.6	6.7	3.9
	DSP-reset	4772.0	3590.0	339.0	13.2	790.0	120.0	13.4	8.5
	MW	19.9	12.0	7.0	**2.3**	8.5	7.4	4.4	**2.1**
	MND	17.1	11.3	7.3	3.9	8.4	7.5	5.5	2.6
	CMC	**14.4**	**10.2**	**6.3**	2.4	**7.5**	**6.6**	**4.2**	**2.1**

We note first that algorithm CPX-Benders generates Benders cuts and also other valid inequalities, in particular when it automatically switches to the standard branch-and-cut solution approach (something not under our control). Table 2 just reports the number of Benders optimality cuts, while the total number of cuts in algorithm CPX-Benders is usually much larger. In some extreme cases, `CPLEX`'s automatic Benders algorithm completely relies on the standard branch-and-cut paradigm. Therefore, we have decided to *not* consider algorithm CPX-Benders when highlighting the algorithm that generates the smallest number of cuts (in bold).

We see that algorithm CMC requires the smallest number of Benders cuts/iterations in all but one case (50%/15, general demand, $a = 200$). Algorithm CMC needs significantly fewer cuts than algorithms MW and MND. For more difficult instances with a larger total number of cuts, the difference sometimes exceeds a factor of two. In summary, the number of generated cuts is closely related to the runtime, see Table 1.

5.5. Comparison of Computational Performance for the CFLP

The *capacitated facility location problem* (CFLP) is fundamental problem in location planning and it has applications in distribution and production planning. It consists in selecting facilities from a finite set of potential facilities and in allocating customer demands in such a way as to minimize fixed costs and transportation costs (Klose and Görtz, 2007). The CFLP and its variations are also suitable for Benders decomposition (Wong, 1978; Wentges, 1996). The Benders master problem selects the facilities to open, and the Benders subproblem decides about the assignment of customers to facilities. We will here focus on the comparison of the performance of Benders algorithms that use different types of Benders cuts, including the new CMCs.

Table 3: Average runtime (in milliseconds) and average number of Benders cuts of the compared approaches for the CFLP. Minima in each group are indicated with bold numbers.

	runtime [ms]			number of cuts		
	number of facilities			number of facilities		
algorithm	$n = 16$	25	50	$n = 16$	25	50
DSP-warm	18.8	86.3	1495.4	19.8	51.00	175.17
MW	23.6	88.7	418.8	**10.8**	**26.3**	49.0
MND	16.4	57.0	255.7	12.4	30.0	52.0
CMC	**16.0**	**49.6**	**230.3**	11.0	26.9	**46.5**

We compare the performance of the four fastest Benders approaches, i.e., algorithms DSP-warm, MW, MND, and CMC. Any other solution approach (directly MIP-based, branch-and-cut, branch-and-price etc.) is not considered here, even if some of these approaches constitute the state of the art. There is a rich literature on the topic (see, e.g., Klose and Görtz, 2007; Fischetti et al., 2016). Appendix D provides the MIP model that serves as the original formulation used for the Benders decomposition. Moreover, the Benders master problem, the subproblem, its dual, and the CMC generation subproblem are presented. Estimates for the weights ε_1 and ε_2 are also given.

We use the 37 CFLP benchmark instances from the OR-Lib (Beasley, 2015), where we group instances by the number n of potential facilities. The computational setup is the same as described in Section 5.3.

Table 3 shows the average runtime and number of iterations of the four Benders algorithms. We see that algorithm CMC clearly outperforms the standard Benders algorithm DSP-warm and the algorithm MW that uses the cuts of Magnanti and Wong (1981). Compared to algorithm MND, the runtime improves moderately when using algorithm CMC, resulting in shortest runtimes in all three groups.

5.6. *Comparison of Computational Performance for the FCNFP*

The *fixed charge network flow problem* (FCNFP) is the problem of routing homogeneous goods at minimum total cost, where each arc of the network has variable cost per unit of flow and a fixed cost which must be paid if any flow uses the arc. The FCNFP has applications in transportation and logistics, telecommunication, and network design. The FCNFP has been solved successfully with Benders decomposition (Costa, 2005). The Benders master program selects the arcs that the flow can use, while the subproblem is a standard network flow problem, which can be solved efficiently with LP or network flow algorithms (Ahuja et al., 1993).

As in the previous section, we compare four Benders implementations: algorithm DSP-warm, MW, MND, and CMC. Note that the FCNFP has been used to exemplify the benefit of using maximum nondominated cuts instead of MW cuts (Sherali and Lunday, 2013). Appendix E presents the FCNFP model, the Benders master problem, the subproblem, the dual subproblem, and the CMC generation subproblem. Moreover, bounds for the weights ε_1 and ε_2 are provided.

We generate instances as described by Sherali and Lunday (2013) using the `C++` code kindly provided by Brian J. Lunday (2025). We generate nine groups of random instances defined by digraphs of identical size with $n = |V|$ vertices and $m = |A|$ arcs, where $n \in \{20, 35, 50\}$ and $m \in \{2n, 2n + 25, 2n + 50\}$. Each group comprise 20 instances, resulting in a benchmark set of size 180. Note that we generated smaller digraphs than those used in the experiments conducted by Sherali and Lunday, because they solved only a small fraction of their instances optimally. The computational setup is again the same as described in Section 5.3.

Table 4: Average runtime (in milliseconds) and average number of Benders cuts of the compared approaches for the FCNFP. Minima in each group are indicated with bold numbers.

		number of vertices								
		$n = 20$			$n = 35$			$n = 50$		
		number of arcs			number of arcs			number of arcs		
	algorithm	$m = 40$	65	90	$m = 70$	95	120	$m = 100$	125	150
runtime [ms]	DSP-warm	28.1	151.2	511.0	36.0	727.8	12065.5	56.5	1935.6	25599.6
	MW	31.9	1038.4	6102.1	41.4	31806.8	194549.2	65.8	10063.9	266803.8
	MND	**24.4**	142.8	391.5	34.1	638.2	13222.3	52.3	1864.4	24237.8
	CMC	**24.4**	**140.6**	**351.0**	**34.0**	**612.3**	**9468.7**	**52.0**	**1738.2**	**20311.7**
number of cuts	DSP-warm	44.0	106.0	171.0	48.6	175.7	336.0	61.9	238.4	411.2
	MW	41.2	445.3	836.2	**46.8**	1090.3	3278.2	61.0	721.5	3025.8
	MND	40.3	104.2	150.0	48.8	177.2	341.4	60.3	217.8	**381.0**
	CMC	**40.4**	**101.4**	**142.8**	48.9	**174.1**	**315.7**	**60.1**	**204.4**	386.5

Table 4 shows the average runtime and number of Benders iterations of the four Benders algorithms. Surprisingly, for a few instances, the algorithm MW generated a very large number of feasibility cuts. In these cases, algorithm MW did not solve the instances within the time limit, leading to strongly distorted average runtimes. These instances also have a dominating effect on the total number of Benders cuts. The comparison of algorithms MND and CMC shows that both can solve all 180 instances well. The Benders algorithm using the new core-maximal cuts slightly outperforms the one using maximum nondominated cuts, with a more substantial reduction in computation time when the number of arcs becomes larger, i.e., for less sparse underlying digraphs.

6. Conclusions

The paper introduced a new solution approach to the SPRP-SS based on Benders decomposition. This approach starts with a new integer programming model, formulation PB, which separates the selection of pick positions from the modeling of the picker route. The Benders subproblem is efficiently solvable via dynamic programming, because it is a shortest path problem defined over the dynamic-programming state space that describes the picker routing problem resulting from solving the Benders master problem. However, a straightforward Benders algorithms is often outperformed by state-of-the-art MIP solver-based approaches, such as those presented by Heßler and Irnich (2024) and Lüke et al. (2025). The main issues are the Benders approach's slow convergency and extensive runtime, as it requires a large number of iterations.

Therefore, we analyzed different dual subproblems to better control the computation of dual optimal solutions and optimality cuts. To this end, Magnanti and Wong (1981) introduced by Pareto-optimal Benders cuts. We analyzed their subproblem and investigated what determines the quality of a strong cut. Based on these insights, we derived a new type of dual subproblem to generate so-called core-maximal cuts. A core-maximal cut aims to maximize the constant part of a Benders cut while preserving dual optimality. Furthermore, it maximizes the coefficients of the master variables in the cut. Our proposed dual subproblem procedure consists of three lexicographic subproblems. For the SPRP-SS application, we showed that these subproblems can be combined into a single dual subproblem that generates optimal CMCs.

In the computational experiments, we compared three MIP solver-based approaches for the SPRP-SS with several Benders decomposition algorithms that differ in how they generate optimality cuts. The latter's cut selection methods include simple versions that rely on the direct solution of subproblems or the use of an LP solver. More sophisticated methods include those of Magnanti and Wong, as well as Sherali and Lunday (2013), which generate maximum nondominated Benders cuts. In summary, the Benders algorithm using the new core-maximal cuts was the fastest in 24 out of 40 groups of instances. The extensive study included 2000 SPRP-SS instances covering a broad range of possibilities for scattering articles in a warehouse. Moreover, of the ten algorithms compared, the Benders approach with the CMCs was the fastest for 870 of the 2000 instances. The analysis via performance profile also confirmed the superiority of the new algorithms.

The relevance of CMCs extends beyond the picker routing application. Two experiments with the CFLP and FCNFP show that the new cut selection method often yields a shorter runtime and fewer Benders iterations. Unlike two-stage stochastic optimization problems that decompose by scenario, all the aforementioned problems are examples, in which the subproblem is not separable. These results confirm that "the Benders approach can be very effective even without separability, as its performance is comparable and sometimes even better than that of the most effective and sophisticated algorithms proposed in the previous literature" (Fischetti et al., 2016, p. 557).

Acknowledgement

We would like to thank Brian J. Lunday (2025) from the Air Force Institute of Technology (AFIT) the for providing his `C++` code to generate FCNFP instances.

This research was supported by Deutsche Forschungsgemeinschaft (DFG) under grant IR 122/13 of project 555303283. We gratefully acknowledge this support.

References

R. Ahuja, T. Magnanti, and J. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Englewood Cliffs, New Jersey, 1993.

J. Beasley. ORLIB. `http://people.brunel.ac.uk/~mastjjb/jeb/orlib/capinfo.html`, 2015. Accessed: 2025-07-21.

J. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4(1):238–252, 1962. doi:10.1007/BF01386316.

N. Boysen, R. de Koster, and F. Weidinger. Warehousing in the e-commerce era: A survey. *European Journal of Operational Research*, 277(2):396–411, 2019. doi:10.1016/j.ejor.2018.08.023.

A. M. Costa. A survey on benders decomposition applied to fixed-charge network design problems. *Computers and Operations Research*, 32:1429–1450, 2005. doi:10.1016/J.COR.2003.11.012.

R. L. Daniels, J. L. Rummel, and R. Schantz. A model for warehouse order picking. *European Journal of Operational Research*, 105(1):1–17, 1998. doi:10.1016/S0377-2217(97)00043-X.

G. B. Dantzig and P. Wolfe. Decomposition principle for linear programs. *Operations Research*, 8(1):101–111, 1960. doi:10.1287/opre.8.1.101.

R. de Koster, T. Le-Duc, and K. J. Roodbergen. Design and control of warehouse order picking: A literature review. *European Journal of Operational Research*, 182(2):481–501, 2007. doi:10.1016/j.ejor.2006.07.009.

E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91 (2):201–213, 2002.

J. Drury. Towards more efficient order picking. Technical report, The Institute of Materials Management, Cranfield, UK, 1988.

M. Fischetti, D. Salvagnin, and A. Zanette. A note on the selection of Benders' cuts. *Mathematical Programming*, 124:175–182, 2010. doi:10.1007/s10107-010-0365-7.

M. Fischetti, I. Ljubić, and M. Sinnl. Benders decomposition without separability: A computational study for capacitated facility location problems. *European Journal of Operational Research*, 253(3):557–569, 2016. doi:10.1016/j.ejor.2016.03.002.

E. Frazelle. *World-Class Warehousing and Material Handling*. McGraw-Hill, New York, 2002.

L. Glomb, F. Liers, and F. Rösel. A novel Pareto-optimal cut selection strategy for Benders decomposition. *Mathematical Programming Computation*, 2025. doi:10.1007/s12532-025-00291-1.

D. Goeke and M. Schneider. Modeling single-picker routing problems in classical and modern warehouses. *INFORMS Journal on Computing*, 33(2):436–451, 2021. doi:10.1287/ijoc.2020.1040.

J. Gu, M. Goetschalckx, and L. F. McGinnis. Research on warehouse operation: A comprehensive review. *European Journal of Operational Research*, 177(1):1–21, 2007. doi:10.1016/j.ejor.2006.02.025.

K. Heßler and S. Irnich. A note on the linearity of Ratliff and Rosenthal's algorithm for optimal picker routing. *Operations Research Letters*, 50(2):155–159, 2022. doi:10.1016/j.orl.2022.01.014.

K. Heßler and S. Irnich. Exact solution of the single picker routing problem with scattered storage. *INFORMS Journal on Computing*, 36(6):1417–1435, 2024. doi:10.1287/ijoc.2023.0075.

A. Klose and S. Görtz. A branch-and-price algorithm for the capacitated facility location problem. *European Journal of Operational Research*, 179(3):1109–1125, 2007. doi:10.1016/j.ejor.2005.03.078.

H. Lin and H. Üster. Exact and heuristic algorithms for data-gathering cluster-based wireless sensor network design problem. *IEEE/ACM Transactions on Networking*, 22(3):903–916, 2013. doi:10.1109/TNET.2013.2262153.

B. J. Lunday. Private communication, 2025. Department of Operational Sciences, Air Force Institute of Technology (AFIT) `https://www.afit.edu/bios/bio.cfm?facID=248`.

L. Lüke, K. Heßler, and S. Irnich. The single picker routing problem with scattered storage: Modeling and evaluation of routing and storage policies. *OR Spectrum*, 46:909–951, 2024. doi:10.1007/s00291-024-00760-4.

L. Lüke, A. Hessenius, and S. Irnich. A linear-size model for the single picker routing problem with scattered storage. *European Journal of Operational Research*, 2025. doi:10.1016/j.ejor.2025.11.002.

T. L. Magnanti and R. T. Wong. Accelerating Benders decomposition: Algorithmic enhancement and model selection criteria. *Operations Research*, 29(3):464–484, 1981. doi:10.1287/opre.29.3.464.

M. Masae, C. H. Glock, and E. H. Grosse. Order picker routing in warehouses: A systematic literature review. *International Journal of Production Economics*, 224:107564, 2020. doi:10.1016/j.ijpe.2019.107564.

C. G. Petersen. An evaluation of order picking routeing policies. *International Journal of Operations & Production Management*, 17(11):1098–1111, 1997. doi:10.1108/01443579710177860.

R. Rahmaniani, T. G. Crainic, M. Gendreau, and W. Rei. The Benders decomposition algorithm: A literature review. *European Journal of Operational Research*, 259(3):801–817, 2017. doi:10.1016/j.ejor.2016.12.005.

G. R. Raidl. Decomposition based hybrid metaheuristics. *European Journal of Operational Research*, 244(1):66–76, 2015. doi:10.1016/j.ejor.2014.12.005.

C. Randazzo, H. P. L. Luna, and P. Mahey. Benders decomposition for local access network design with two technologies. *Discrete Mathematics & Theoretical Computer Science*, 4, 2001. doi:10.46298/dmtcs.288.

H. D. Ratliff and A. S. Rosenthal. Order-picking in a rectangular warehouse: A solvable case of the traveling salesman problem. *Operations Research*, 31(3):507–521, 1983. doi:10.1287/opre.31.3.507.

R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970. ISBN 978-0691015866.

K. J. Roodbergen and R. de Koster. Routing order pickers in a warehouse with a middle aisle. *European Journal of Operational Research*, 133(1):32–43, 2001. doi:10.1016/s0377-2217(00)00177-6.

G. K. Saharidis, M. Minoux, and M. G. Ierapetritou. Accelerating Benders method using covering cut bundle generation. *International Transactions in Operational Research*, 17(2):221–237, 2010. doi:10.1111/j.1475-3995.2009.00706.x.

H. D. Sherali and B. J. Lunday. On generating maximal nondominated Benders cuts. *Annals of Operations Research*, 210: 57–72, 2013. doi:10.1007/s10479-011-0883-6.

H. D. Sherali and A. L. Soyster. Preemptive and nonpreemptive multi-objective programming: Relationship and counterexamples. *Journal of Optimization Theory and Applications*, 39(2):173–186, 1983. doi:10.1007/bf00934527.

K. N. Singh and D. L. van Oudheusden. A branch and bound algorithm for the traveling purchaser problem. *European Journal of Operational Research*, 97(3):571–579, 1997. doi:10.1016/S0377-2217(96)00313-X.

P. M. Stursberg. *On the Mathematics of Energy System Optimization: Network Models, Decomposition, and Economic Incentives*. PhD thesis, Technische Universität München, 2019. URL `https://mediatum.ub.tum.de/1476849`.

J. Tompkins, J. White, Y. Bozer, E. Frazelle, and J. Tanchoco. *Facilities Planning*. John Wiley & Sons, Hoboken, NJ, 2003.

T. van Gils, K. Ramaekers, A. Caris, and R. B. de Koster. Designing efficient order picking systems by combining planning problems: State-of-the-art classification and review. *European Journal of Operational Research*, 267(1):1–15, 2018. doi:10.1016/j.ejor.2017.09.002.

F. Weidinger. Picker routing in rectangular mixed shelves warehouses. *Computers & Operations Research*, 95:139–150, 2018. doi:10.1016/j.cor.2018.03.012.

P. Wentges. Accelerating Benders' decomposition for the capacitated facility location problem. *Mathematical Methods of Operations Research*, 44(2):267–290, 1996. doi:10.1007/bf01194335.

R. T. Wong. *Accelerating Benders decomposition for network design*. PhD thesis, Massachusetts Institute of Technology, 1978.

Appendix A. Pareto-Optimality of Magnanti and Wong's Subproblem

Specifically for the SPRP-SS, we show that an optimal solution $(\boldsymbol{\mu}^*, \boldsymbol{\nu}^*)$ to the linear program (12), i.e., a MW optimality cut, is always Pareto-optimal. The proof is by contradiction. Hence, let another dual feasible solution $(\boldsymbol{\mu}', \boldsymbol{\nu}')$ dominate $(\boldsymbol{\mu}^*, \boldsymbol{\nu}^*)$, written out as

$$\mu'_o - \bar{\mu}'_d + \sum_{p \in P} \bar{\nu}'_p (1 - y_p) \geq \mu^\star_o - \mu^\star_d + \sum_{p \in P} \nu^\star_p (1 - y_p) \qquad \forall \boldsymbol{y} \in Y \tag{A.1a}$$

with strict $>$ at least once. We assume $>$ for $\dot{\boldsymbol{y}} \in Y$, i.e.,

$$\mu'_o - \bar{\mu}'_d + \sum_{p \in P} \bar{\nu}'_p (1 - \dot{y}_p) > \mu^\star_o - \mu^\star_d + \sum_{p \in P} \nu^\star_p (1 - \dot{y}_p). \tag{A.1b}$$

Moreover, any point $\boldsymbol{w} \in Y^c$ can be expressed as a convex combination of extreme points. Since $Y \subset \{0,1\}^P$, all $\boldsymbol{y} \in Y$ are these extreme points. Hence, for $\boldsymbol{w} \in Y^c$, there exist weights $\lambda^1, \lambda^2, \ldots, \lambda^k \geq 0$, $\sum_{j=1}^k \lambda^j = 1$ and extreme points $\boldsymbol{y}^1, \boldsymbol{y}^2, \ldots, \boldsymbol{y}^k \in Y$ with $\boldsymbol{w} = \sum_{j=1}^k \lambda^j \boldsymbol{y}^j$. Therefore, linearity and (A.1a) imply

$$\mu'_o - \bar{\mu}'_d + \sum_{p \in P} \bar{\nu}'_p (1 - w_p) \geq \mu^\star_o - \mu^\star_d + \sum_{p \in P} \nu^\star_p (1 - w_p) \qquad \forall \boldsymbol{w} \in Y^c. \tag{A.1c}$$

In particular, the MW point $\tilde{\boldsymbol{y}} \in \text{relint}(Y^c) \subseteq Y^c$ must fulfill

$$\mu'_o - \bar{\mu}'_d + \sum_{p \in P} \bar{\nu}'_p (1 - \tilde{y}_p) = \mu^\star_o - \mu^\star_d + \sum_{p \in P} \nu^\star_p (1 - \tilde{y}_p), \tag{A.1d}$$

where equality results because the value on the RHS is maximal for the MW subproblem solution. Hence, $(\boldsymbol{\mu}', \boldsymbol{\nu}')$ is necessarily also an optimal solution to the MW subproblem (12).

A classical result of convex analysis (Rockafellar, 1970) is that for $\tilde{\boldsymbol{y}} \in \text{relint}(Y^c)$, there must exist a scalar $\psi > 1$ such that $\boldsymbol{w} := \psi \tilde{\boldsymbol{y}} + (1 - \psi)\dot{\boldsymbol{y}} \in Y^c$. Also this point $\boldsymbol{w}$ fulfills (A.1c). Multiplying (A.1d) with ψ and (A.1b) with $1 - \psi < 0$, and adding the results yields

$$\mu'_o - \bar{\mu}'_d + \sum_{p \in P} \bar{\nu}'_p (1 - w_p) < \mu^\star_o - \mu^\star_d + \sum_{p \in P} \nu^\star_p (1 - w_p), \tag{A.1e}$$

contradicting with (A.1c).

Appendix B. Pareto-Optimality of the Core-Maximal Cuts in the SPRP-SS

For the SPRP-SS, we provide insights into the quality of core-maximal cuts in terms of Pareto-optimality. While CMCs are not necessarily Pareto-optimal, we can characterize when they are, and when they are dominated.

Let a core-maximal cut be defined by the dual optimal solution $(\boldsymbol{\mu}, \boldsymbol{\nu})$. We assume that this core-maximal cut is Pareto-dominated. Then there exists a dual optimal solution $(\boldsymbol{\mu}', \boldsymbol{\nu}')$ with $\mu_o - \mu_d + \sum_{p \in P}(1 - y_p)\nu_p \leq \mu'_o - \mu'_d + \sum_{p \in P}(1 - y_p)\nu'_p$ for all $\boldsymbol{y} \in Y$ and strict inequality for at least one $\boldsymbol{y}^* \in Y$. We distinguish three cases:

1. $\mu_o - \mu_d > \mu'_o - \mu'_d$: Since $\boldsymbol{y} = \mathbf{1} = (1)_{p\in P} \in Y$, the CMCs is not dominated, contradicting the assumption.

2. $\mu_o - \mu_d = \mu'_o - \mu'_d$: The maximality of the constant in the CMC implies $\sum_{p\in P} \nu_p \geq \sum_{p\in P} \nu'_p$. Therefore, there must exist a position $p^+ \in P$ with $\nu_{p^+} > \nu'_{p^+}$. We can now define $\hat{\boldsymbol{y}} \in Y$ with $\hat{y}_{p^+} = 0$ and $\hat{y}_p = 1$ for all $p \neq p^+$. It follows that $\sum_{p\in P}(1-\hat{y}_p)\nu_p = \nu_{p^+} > \nu'_{p^+} = \sum_{p\in P}(1-\hat{y}_p)\nu'_p$. Thus, the core-maximal cut is not dominated, contradicting the assumption.

3. $\mu_o - \mu_d < \mu'_o - \mu'_d$: Dual optimality of $(\boldsymbol{\mu}, \boldsymbol{\nu})$ and $(\boldsymbol{\mu}', \boldsymbol{\nu}')$ implies

$$z_{\mathrm{SP}(\bar{y})} = \mu_o - \mu_d + \sum_{p\in P^0(\boldsymbol{y})} \nu_p = \mu'_o - \mu'_d + \sum_{p\in P^0(\boldsymbol{y})} \nu'_p.$$

Let $\delta := (\mu'_o - \mu'_d) - (\mu_o - \mu_d) > 0$. Then, $\delta = \sum_{p\in P^0(\boldsymbol{y})}(\nu_p - \nu'_p)$.

Moreover, define $\Delta := \sum_{p\in P^1(\bar{\boldsymbol{y}})}(\nu_p - \nu'_p)$. Then, $\Delta > 0$ because $(\boldsymbol{\mu}, \boldsymbol{\nu})$ defines a core-maximal cut. Summarizing, the only possibility for a core-maximal cut to not be Pareto-optimal is when $\mu_o - \mu_d$ is not maximal compared to other dual optimal cuts.

The further analysis is involved. Therefore, we consider only the simplified case that $\nu_{p_0} - \nu'_{p_0} = \delta$ and $\nu_{p_1} - \nu'_{p_1} = \Delta$ for two positions $p_0, p_1 \in P$ (the analysis can be extended to sets of positions P_0 and P_1). For an arbitrary master solutions $\bar{\boldsymbol{y}} \in Y$, consider the following four cases regarding the alterative optimality cut derived from $(\boldsymbol{\mu}', \boldsymbol{\nu}')$:

(a) $\bar{y}_{p_0} = \bar{y}_{p_1} = 1$: It yields a higher bound differing by δ.

(b) $\bar{y}_{p_0} = 1, \bar{y}_{p_1} = 0$: It yields a bound differing by $\delta - \Delta$.

(c) $\bar{y}_{p_0} = 0, \bar{y}_{p_1} = 1$: It yields the identical bound.

(d) $\bar{y}_{p_0} = \bar{y}_{p_1} = 0$: It yields a worse bound differing by Δ.

However, if no feasible solution with $\bar{y}_{p_0} = \bar{y}_{p_1} = 0$ exists and $\Delta \leq \delta$, then a cut derived from $(\boldsymbol{\mu}, \boldsymbol{\nu})$ is dominated by a cut derived from $(\boldsymbol{\mu}', \boldsymbol{\nu}')$. Therefore, we cannot guarantee Pareto-optimal cuts in the sense of Magnanti and Wong.

Appendix C. Example of the Unboundedness of the MW Subproblem

We show that the MW subproblem can be unbounded. Consider the following example:

$$\begin{aligned} \min \quad & 0 \\ \text{subject to} \quad & x_1 \geq 1 \\ & -x_1 \geq -1 + 2y \\ & y \in \{0,1\} \\ & x_1, x_2 \geq 0 \end{aligned}$$

Here, we decompose into $y \in Y = \{0,1\}$ and $(x_1, x_2) \in \mathbb{R}^2$. The Benders master problem, the Benders subproblem, and its dual are then given by:

Master problem:

$$\begin{aligned} \min \quad & z \\ \text{subject to} \quad & \{\text{cuts}\} \\ & y \in \{0,1\} \end{aligned}$$

Subproblem:

$$\begin{aligned} \min \quad & 0 \\ \text{subject to} \quad & x_1 \geq 1 \\ & -x_1 \geq -1 + 2\bar{y} \\ & x_1, x_2 \geq 0 \end{aligned}$$

Dual subproblem:

$$\begin{aligned} z_{\text{DSP}(\bar{y})} = \max \quad & \pi_1 + \pi_2(-1 + 2\bar{y}) \\ \text{subject to} \quad & \pi_1 - \pi_2 \leq 0 \\ & \pi_1, \pi_2 \geq 0 \end{aligned}$$

For $\bar{y} = 0 \in Y$, the DSP objective becomes $z_{\text{DSP}(0)} = \pi_1 - \pi_2$, which is optimal for $\pi_1 = \pi_2 = 0$ with $z_{\text{DSP}(0)} = 0$.

For $y \in \{0,1\} = Y$ and a point in the (relative) interior of Y, i.e., $\tilde{y} \in (0,1) = \text{relint}(Y^c)$, the MW subproblem is:

$$\begin{aligned} z_{\text{MW}(\bar{y},\tilde{y})} = \max \quad & \pi_1 + \pi_2(-1 + 2\tilde{y}) \\ \text{subject to} \quad & \pi_1 - \pi_2 \leq 0 \\ & \pi_1 + \pi_2(-1 + 2\bar{y}) = z_{\text{DSP}}(\bar{y}) \\ & \pi_1, \pi_2 \geq 0 \end{aligned}$$

For $\tilde{y} = 0.5 \in \text{relint}(Y^c)$, which is feasible for the Benders master, the MW objective becomes $z_{\text{MW}(\bar{y},0.5)} = \pi_1$ independent of $\bar{y}$. For $\bar{y} = 0$, the MW constraint is $\pi_1 - \pi_2 = 0$, i.e., $\pi_1 = \pi_2$ describing the entire domain of the DSP. Hence, the DSP is unbounded.

In summary, we showed that the MW subproblem can be unbounded, even though the DSP is bounded. The same is true for the new subproblem (17) used to generate core-maximal cuts. If the constant in the optimality cut can become arbitrarily large, then the first auxiliary problem (15) is unbounded.

Appendix D. Capacitied Facility Location Problem

The CFLP can be defined as follows. We are given a set J of facilities, each with a capacity s_j and fixed cost f_j for $j \in J$. Moreover, a set I of customers is given, each with a demand d_i for $i \in I$. For each pair $(i,j) \in I \times J$, the cost coefficient c_{ij} is the cost of transporting one unit of demand from facility j to customer i. The goal is to open facilities and assign customers to the open facilities to minimize the total cost, comprising fixed cost incurred by the opened facilities and the variable cost resulting from the assignment.

We use the classical CFLP formulation as presented by Fischetti et al. (2016). It uses binary variables $w_j \in \{0,1\}$ for $j \in J$ to indicate whether facility j is opened, and continuous variables $0 \leq x_{ij} \leq 1$ for

$(i, j) \in I \times J$, which represent the fraction of the total demand of customer i that is covered by facility j.

$$z_{\text{CFLP}} = \min \sum_{j \in J} f_j w_j + \sum_{i \in I} \sum_{j \in J} c_{ij} d_i x_{ij} \tag{D.1a}$$

$$\text{subject to} \quad \sum_{j \in J} x_{ij} = 1 \qquad \forall i \in I \tag{D.1b}$$

$$\sum_{i \in I} d_i x_{ij} \leq s_j w_j \qquad \forall j \in J \tag{D.1c}$$

$$x_{ij} \leq w_j \qquad \forall i \in I, j \in J \tag{D.1d}$$

$$\sum_{j \in J} s_j w_j \geq \sum_{i \in I} d_i \tag{D.1e}$$

$$x_{ij} \geq 0 \qquad \forall i \in I,\ j \in J \tag{D.1f}$$

$$w_j \in \{0, 1\} \qquad \forall j \in J \tag{D.1g}$$

The objective (D.1a) is to minimize the total cost. Constraints (D.1b) ensure that each customer is covered. Capacity constraints are given by (D.1c). The direct coupling of location and allocation decisions is established by (D.1d), which are redundant constraints. The coverage of the total customer demand by the opened locations is ensured by (D.1e). Lastly, (D.1f) and (D.1g) define the variables' domains.

To account for the desired form of the formulation regarding the maximization of the constant part of the optimality cut (see Section 4), we substitute the binary variables w_j by the complementary binary variables $y_j = 1 - w_j$ for all $j \in J$. Then, the constraints (D.1e) become

$$\sum_{j \in J} s_j y_j \leq \sum_{j \in J} s_j - \sum_{i \in I} d_i \qquad \forall j \in J. \tag{D.2}$$

The straightforward Benders decomposition with $Y = \{\boldsymbol{y} \in \{0, 1\}^J : (\text{D.2})\}$ is:

Master problem:

$$z_{\text{BM}} = \min \sum_{j \in J} f_j (1 - y_j) + z \tag{D.3a}$$

$$\text{subject to} \quad \{\text{opt. cuts}\} \tag{D.3b}$$

$$\boldsymbol{y} \in Y, z \geq 0 \tag{D.3c}$$

Subproblem:

$$z_{\text{SP}(\bar{\boldsymbol{y}})} = \min \sum_{i \in I} \sum_{j \in J} c_{ij} d_i x_{ij} \tag{D.4a}$$

$$\text{subject to} \quad \sum_{j \in J} x_{ij} = 1 \qquad \forall i \in I \tag{D.4b}$$

$$\sum_{i \in I} d_i x_{ij} \leq s_j (1 - \bar{y}_j) \qquad \forall j \in J \tag{D.4c}$$

$$x_{ij} \leq 1 - \bar{y}_j \qquad \forall i \in I, j \in J \tag{D.4d}$$

$$x_{ij} \geq 0 \qquad \forall i \in I, j \in J \tag{D.4e}$$

This subproblem SP($\bar{\boldsymbol{y}}$) is always feasible because of (D.2), which avoids feasibility cuts in Benders' algorithm. SP($\bar{\boldsymbol{y}}$) tends to have a higher degree of degeneracy because of the redundant constraints (D.4d) equivalent to (D.1d). This fact distinguishes it from subproblems of the simple/uncapacitated facility location problem and CFLP models without the redundant constraints (D.4d), where the procedure to generate Benders cuts does not have a significant impact.

Given model (D.4) of the Benders subproblem, the dual subproblem DSP($\bar{\boldsymbol{y}}$) has three types of variables:

$\boldsymbol{\pi} = (\pi_i)_{i\in I} \in \mathbb{R}^I$ for the constraints (D.4b), $\boldsymbol{\mu} = (\mu_j)_{j\in J} \leq \mathbf{0}$ for the constraints (D.4c), and $\boldsymbol{\nu} = (\nu_{ij})_{(i,j)\in I\times J} \leq \mathbf{0}$ for the constraints (D.4d). The formulation is:

$$z_{\mathrm{DSP}(\bar{\boldsymbol{y}})} = \max \sum_{i\in I} \pi_i + \sum_{j\in J} s_j(1-\bar{y}_j)\mu_j + \sum_{i\in I}\sum_{j\in J}(1-\bar{y}_j)\nu_{ij} \tag{D.5a}$$

$$\text{subject to} \quad \pi_i + d_i\mu_j + \nu_{ij} \leq d_i c_{ij} \qquad \forall i \in I, j \in J \tag{D.5b}$$

$$\boldsymbol{\pi} \in \mathbb{R}^I, \quad \boldsymbol{\mu} \leq \mathbf{0}, \quad \boldsymbol{\nu} \leq \mathbf{0} \tag{D.5c}$$

For a feasible dual solution $(\bar{\boldsymbol{\pi}}, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}})$, the optimality cuts is

$$z \geq \sum_{i\in I} \bar{\pi}_i + \sum_{j\in J} s_j(1-y_j)\bar{\mu}_j + \sum_{i\in I}\sum_{j\in J}(1-y_j)\bar{\nu}_{ij},$$

where the right-hand side of the optimality cut has the constant part and variable part

$$\mathrm{const}(\bar{\boldsymbol{\pi}}, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}}) := \sum_{i\in I} \bar{\pi}_i + \sum_{j\in J} s_j\bar{\mu}_j + \sum_{i\in I}\sum_{j\in J} \bar{\nu}_{ij} \quad \text{and} \quad \mathrm{var}(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}}) := -\sum_{j\in J} s_j\bar{\mu}_j y_j - \sum_{i\in I}\sum_{j\in J} \bar{\nu}_{ij} y_j, \tag{D.6a}$$

respectively. The subproblem to generate core-maximal cuts is the following:

$$z_{\mathrm{CMC}(\bar{\boldsymbol{y}})} = \max \sum_{i\in I}(1+\varepsilon_1)\pi_i + \sum_{j\in J} s_j(1+\varepsilon_1-\varepsilon_2-\bar{y}_j)\mu_j + \sum_{i\in I, j\in J}(1+\varepsilon_1-\varepsilon_2-\bar{y}_j)\nu_{ij} \tag{D.7a}$$

$$= \max(1+\varepsilon_1)\left(\sum_{i\in I}\pi_i + \sum_{\substack{j\in J:\\ \bar{y}_j=0}} s_j\mu_j + \sum_{\substack{i\in I, j\in J:\\ \bar{y}_j=0}} \nu_{ij}\right) - \varepsilon_2\left(\sum_{\substack{j\in J:\\ \bar{y}_j=0}} s_j\mu_j + \sum_{\substack{i\in I, j\in J:\\ \bar{y}_j=0}} \nu_{ij}\right) \tag{D.7b}$$

$$+ (\varepsilon_1-\varepsilon_2)\left(\sum_{\substack{j\in J:\\ \bar{y}_j=1}} s_j\mu_j + \sum_{\substack{i\in I, j\in J:\\ \bar{y}_j=1}} \nu_{ij}\right)$$

$$\text{subject to} \quad \text{(D.5b) and (D.5c)}, \tag{D.7c}$$

We can rewrite the objective (D.7a) in a similar fashion as in (18b) to derive bounds for ε_1 and ε_2. For an FCNFP instance, let B be a bound to the maximum variation in $\sum_{j\in J} s_j\mu_j + \sum_{i\in I, j\in J}\nu_{ij}$, which exists due to $\mu_j, \nu_{ij} \in [-\max_{i\in I, j\in J} d_i c_{ij}, 0]$ for all $i \in I, j \in J$. The choice of ε_1 and ε_2 must result in a dual optimal cut with the maximum constant and, regarding this constant, a maximal sum of coefficients for the $\boldsymbol{y}$-variables.

1. First, model (D.7) has to ensure dual optimality. Using similar arguments as for the SPRP-SS, a minimum gain of $1+\varepsilon_1$ in the first term of the objective (D.7b) comes with a maximum change of $\varepsilon_1 B$ in the second and third term. For that reason, we choose $\varepsilon_1 \leq 1/(B-1)$.

2. Given a dual optimal solution, the first term in (D.7b) is fixed. To maximize the constant, the largest change in the second term, which is bounded by $\varepsilon_2 B$, has to be smaller than the minimum change in the third term, which is given by $\varepsilon_1 - \varepsilon_2$. For a dual optimal solution $(\bar{\boldsymbol{\pi}}, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}})$ and $\varepsilon_2 < \varepsilon_1/(B-1)$, model (D.7) maximizes the third term, since $\mathrm{const}(\bar{\boldsymbol{\pi}}, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}}) = z_{\mathrm{CMC}(\bar{\boldsymbol{y}})} + \sum_{\substack{j\in J:\\ y_j=1}} s_j\bar{\mu}_j + \sum_{i\in I}\sum_{\substack{j\in J:\\ y_j=1}} \bar{\nu}_{ij}$.

3. Finally, the second term itself trivially maximizes the coefficients of the $\boldsymbol{y}$-variables given a dual optimal solution with a maximal constant.

4. Unlike in the SPRP-SS, for the CFLP, we need to consider the fact that the constant of an optimality cut can be unbounded when derived from (D.5). Since the generation of core-maximal cuts leads to a perturbation of the subproblem coefficients, extreme rays that were previously neutral for the objective function but had a positive contribution to the constant can now lead to unboundedness of the core-maximal cut-subproblem. Therefore, we need to make sure that the additional gain $\varepsilon_1 \sum_{i\in I} d_i$ is less than the additional cost $(\varepsilon_1 - \varepsilon_2)\sum_{j\in J} s_j$. Together with previously derived bound on ε_2, this leads to $(\varepsilon_1 - \varepsilon_2) \geq \max\{(1+\varepsilon_1)(\sum_{i\in I} d_i - \sum_{j\in J} \bar{y}_j s_j)/(\sum_{j\in J} s_j), B|J|\varepsilon_2\}$. Note that due to the bound on ε_1, unboundedness due to coefficient perturbation can only occur if $\sum_{i\in I} d_i = \sum_{j\in J}(1-\bar{y}_j)s_j$.

Appendix E. Fixed Charge Network Flow Problem

The FCNFP is defined by a directed graph $G = (V, A)$ with capacities $u_{ij} \in \mathbb{Z}_+$, variable costs $c_{ij} \in \mathbb{Z}_+$, and fixed costs $f_{ij} \in \mathbb{Z}_+$ for all arcs $(i, j) \in A$, and supply/demand quantities $d_i \in \mathbb{Z}$ for all vertices $i \in V$.

The classical model of the FCNFP uses two types of variables: continuous flow variables $x_{ij} \geq 0$ and binary arc selection variables $w_{ij} \in \{0, 1\}$ for all arcs $(i, j) \in A$. The model is:

$$z_{\text{FCNFP}} = \min \sum_{(i,j)\in A} (f_{ij} w_{ij} + c_{ij} x_{ij}) \tag{E.1a}$$

$$\text{subject to} \quad \sum_{j\in V:(i,j)\in A} x_{ij} - \sum_{j\in V:(j,i)\in A} x_{ij} = d_i \qquad \forall i \in V \tag{E.1b}$$

$$x_{ij} \leq u_{ij} w_{ij} \qquad \forall (i,j) \in A \tag{E.1c}$$

$$w_{ij} \in \{0,1\}, \quad x_{ij} \geq 0 \qquad \forall (i,j) \in A \tag{E.1d}$$

The objective (E.1a) is to minimize the sum of the variable and fixed costs. Constraints (E.1b) ensure flow conservation, and (E.1c) couple flow variables and selection variables. The domains of the variables are specified in (E.1d).

For the Benders decomposition, we again substitute the $\boldsymbol{w}$-variables by complementary variables, i.e., variables $\boldsymbol{y}$ defined by $y_{ij} = 1 - w_{ij}$ for all $(i, j) \in A$. Now, we can define $Y = \{0, 1\}^A$ so that the Benders decomposition is:

Master problem:

$$z_{\text{BM}} = \min \sum_{(i,j)\in A} f_{ij}(1 - y_{ij}) + z \tag{E.2a}$$

$$\text{subject to} \quad \{\text{opt. cuts}\} \tag{E.2b}$$

$$\{\text{feas. cuts}\} \tag{E.2c}$$

$$\boldsymbol{y} \in Y, z \geq 0 \tag{E.2d}$$

Subproblem:

$$z_{\text{SP}(\bar{\boldsymbol{y}})} = \min \sum_{(i,j)\in A} c_{ij} x_{ij} \tag{E.3a}$$

$$\text{subject to} \quad \mathcal{N}\boldsymbol{x} = \boldsymbol{d} \tag{E.3b}$$

$$x_{ij} \leq u_{ij}(1 - \bar{y}_{ij}) \qquad \forall (i,j) \in A \tag{E.3c}$$

$$\boldsymbol{x} \geq \boldsymbol{0} \tag{E.3d}$$

In the Benders subproblem (E.3), the flow variables are aggregated into $\boldsymbol{x} = (x_{ij})_{(i,j)\in A} \geq \boldsymbol{0}$ and the supply/demand values are aggregated into $\boldsymbol{d} = (d_i)_{i\in V}$. Recall that $\mathcal{N}$ describes the incidence matrix of the digraph (V, A).

The master problem is given by MIP (E.2) and the subproblem is shown in LP (E.3). The variables in the dual subproblem DSP($\bar{\boldsymbol{y}}$) are $\boldsymbol{\mu} = (\mu_i)_{i\in V} \in \mathbb{R}^V$ for the constraints (E.3b) and $\boldsymbol{\nu} = (\nu_{ij})_{(i,j)\in A} \leq \mathbf{0}$ for constraints (E.3c). The following dual subproblem results:

$$z_{\mathrm{DSP}(\bar{\boldsymbol{y}})} = \max \sum_{i\in V} d_i \mu_i + \sum_{(i,j)\in A} u_{ij}(1-\bar{y}_{ij})\nu_{ij} \tag{E.4a}$$

$$\text{subject to} \quad \mu_i - \mu_j + \nu_{ij} \leq c_{ij} \qquad \forall (i,j)\in A \tag{E.4b}$$

$$\boldsymbol{\mu} \in \mathbb{R}^V, \quad \boldsymbol{\nu} \leq \mathbf{0} \tag{E.4c}$$

Note that the master problem (E.2) does not guarantee a feasible solution. Therefore, feasibility cuts need to be generated. For a unbounded dual direction $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}})$, they have the form

$$0 \geq \sum_{i\in V} d_i\bar{\mu}_i + \sum_{(i,j)\in A} u_{ij}(1-y_{ij})\bar{\nu}_{ij}.$$

In the bounded case, the soultion $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}})$ yields the optimality cut

$$z \geq \sum_{i\in V} d_i\bar{\mu}_i + \sum_{(i,j)\in A} u_{ij}(1-y_{ij})\bar{\nu}_{ij}, \tag{E.5a}$$

where the right-hand side of the optimality cut has the constant part and variable part

$$\mathrm{const}(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}}) := \sum_{i\in V} d_i\bar{\mu}_i + \sum_{(i,j)\in A} u_{ij}\bar{\nu}_{ij} \qquad \text{and} \qquad \mathrm{var}(\bar{\boldsymbol{\nu}}) := -\sum_{(i,j)\in A} u_{ij}\bar{\nu}_{ij}y_{ij}, \tag{E.5b}$$

respectively. Lastly, the CMC subproblem is given by:

$$z_{\mathrm{CMC}(\bar{\boldsymbol{y}})} = \max \sum_{i\in V} (1+\varepsilon_1) d_i\mu_i + \sum_{(i,j)\in A} u_{ij}(1+\varepsilon_1-\varepsilon_2-\bar{y}_{ij})\nu_{ij} \tag{E.6a}$$

$$= \max (1+\varepsilon_1)\left(\sum_{i\in V} d_i\mu_i + \sum_{\substack{(i,j)\in A:\\ \bar{y}_{ij}=0}} u_{ij}\nu_{ij}\right) - \varepsilon_2 \left(\sum_{\substack{(i,j)\in A:\\ \bar{y}_{ij}=0}} u_{ij}\nu_{ij}\right) + (\varepsilon_1-\varepsilon_2)\left(\sum_{\substack{(i,j)\in A:\\ \bar{y}_{ij}=1}} u_{ij}\nu_{ij}\right) \tag{E.6b}$$

$$\text{subject to} \quad \text{(E.4b) and (E.4b)} \tag{E.6c}$$

Using a similar idea as for (18b), the objective (E.6a) is rewritten into the equivalent form (E.6b). Let B be an upper bound to the maximum variation in $\sum_{(i,j)\in A} u_{ij}\nu_{ij}$. This bound exists, since $\nu_{ij} \in [-R, 0]$ for all $(i,j)\in A$, where R denotes the length of a longest route in G. We want the CMC subproblem to result in a dual optimal solution with the largest constant and, given the constant, the largest coefficients for the $\boldsymbol{y}$-variables. Thus, we can derive the following bounds for ε_1 and ε_2:

1. First, model (E.6) has to ensure dual optimality. The maximum variation in the second and third term (bounded by B) of the objective (E.6b) has to be smaller than the minimum variation in the first term, which is $1+\varepsilon_1$. This implies $\varepsilon_1 \leq 1/(B-1)$.

2. Given a dual optimal solution, the first term in (E.6b) is fixed. To maximize the constant of the optimality cut, the largest variation in the second term, which is bounded by $\varepsilon_2 B$, has to be smaller than the minimum variation in the third term, which is given by $\varepsilon_1 - \varepsilon_2$. For a dual optimal solution and $\varepsilon_2 < \varepsilon_1/(B-1)$, model (D.7) maximizes the third term since $\text{const}(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}}) = z_{\text{CMC}(\bar{\boldsymbol{y}})} + \sum_{(i,j)\in A:\bar{y}_{ij}=1} u_{ij}\bar{\nu}_{ij}$.
3. Finally, the second term itself trivially maximizes the coefficients of $\boldsymbol{y}$ given a dual optimal solution with a maximal constant.
4. Note that, again, a perturbation of the objective function coefficients can lead to now profitable extreme rays that were not profitable before. However, this can only occur if, for a given subset $S \subset V$, the total capacity of the cut set $\delta^-(S) = \{(i,j) \in A : i \notin S, j \in S\}$ equals the total demand of the vertices in S. In such a case it would be beneficial to increase every $\boldsymbol{\nu}$-variable that correspond to an arc in the cut set, since we can then also increase every $\boldsymbol{\mu}$-variable that corresponds to a vertex in the subset. For these rather specific instances, we set $\varepsilon_2 = \varepsilon_1$ and the CMC subproblem reduces to the regular dual subproblem.