Benders decomposition for choice-based optimization problems with discrete upper-level variables

Stefano Bortolomiol
Virginie Lurkin
Michel Bierlaire
Claudia Bongiovanni

Transport and Mobility Laboratory, EPFL

September 2021

STRC | 21st Swiss Transport Research Conference
Monte Verità / Ascona, September 12 – 14, 2021

Transport and Mobility Laboratory, EPFL

Benders decomposition for choice-based optimization problems with discrete upper-level variables

September 2021

Abstract

In this work, we consider a class of choice-based optimization problems in which the decision variables of the supplier are discrete. First, we compare this class of problems with a more general formulation which admits both continuous and discrete variables. A feasible solution of the problem with both continuous and discrete upper-level variables can be found by solving a problem with only discrete variables by discretizing all continuous variables of the original formulation. An appropriate discretization of all continuous variables can guarantee a good approximation of the solution of the original problem. A computational analysis shows that the discrete formulation is faster than the continuous-discrete formulation for complex problems. Then, we show that in the discrete formulation, in which all utility functions of the customers can be expressed as parameters of the supplier's optimization problem, the lower-level optimization problem of the customer is a continuous knapsack problem. This property is used to derive a Benders decomposition algorithm for the choice-based optimization problem with discrete variables.

Keywords

discrete choice modeling, choice-based optimization, mixed integer optimization, Benders decomposition

1 Introduction

Discrete choice models constitute a state-of-the-art approach to model demand at a disaggregate level, since they account for product differentiation and consumer behavioral heterogeneity. A growing number of works investigate the problem of including discrete choice models into optimization problems. The most used discrete choice models in the optimization literature are the multinomial logit and the nested logit. These models are particularly convenient because they allow to express choice probabilities through a closed-form expression which can be easily incorporated in a nonlinear optimization model. On the contrary, other more advanced discrete choice models have more complex forms which require several simplifying assumptions in order to obtain a tractable formulation.

Discrete choice models can also be integrated into optimization models by means of simulation of the utility function. This approach is outlined in the recent contribution by Pacheco *et al.* (2021), and is applied in a market equilibrium and regulation context by Bortolomiol *et al.* (2021b) and Bortolomiol *et al.* (2021a). A simulation-based approach has the advantage of accommodating a large variety of advanced choice models available in the literature. These complex choice models allow for increasingly complex and precise representations of individual behavior. However, the resulting optimization models suffer from computational limitations which make large-scale instances intractable.

In this work, we start from the framework proposed by Pacheco *et al.* (2021) and we examine some computational aspects of a simple yet general class of choice-based optimization problems in which all variables on the supply side are discrete. In Section 2 we compare a model featuring both continuous and discrete variables with one that has only discrete variables at the supply level. Computational experiments are carried out to evaluate the trade-offs between the two approaches.

This contribution is structured as follows. Section 2 compares a generic model for the choice-based optimization problem with a reformulation which requires all decision variables of the supplier to be discrete and finite. Computational experiments are carried out to evaluate the trade-offs between the two approaches. Section 3 introduces the Benders decomposition algorithm and discusses variants, enhancements and applications. Section 4 discusses how Benders decomposition can be used to solve a particular class of choice-based optimization problems where all supplier variables are discrete and finite. Finally, Section 5 summarizes the main findings and presents the future extensions of this work.

2 Choice-based optimization with continuous and discrete upper-level variables

We consider a market where a discrete and finite set of products are offered to a population. Let N represent the set of customers (or groups of homogeneous customers with size θ_n), who are assumed to be utility maximizers, and let I indicate the set of alternatives available in the market. Utility functions U_{in} are defined for each $n \in N$ and alternative $i \in I$. Each utility function takes into account the socioeconomic characteristics and the tastes of the individual as well as the attributes of the alternative. According to random utility theory (Manski, 1977), U_{in} can be decomposed into a systematic component V_{in} which includes all that is observed by the analyst and a random term ε_{in} which captures the uncertainties caused by unobserved attributes and unobserved taste variations. Therefore, the resulting discrete choice models are naturally probabilistic. The probability that customer n chooses alternative i is defined as

$$P_{in} = \Pr[V_{in} + \varepsilon_{in} = \max_{j \in I}(V_{jn} + \varepsilon_{jn})]. \quad (1)$$

Following the approach introduced by Pacheco *et al.* (2021), we approximate the choice probabilities using a simulation-based linearization. Specifically, a set R of independent draws are extracted from the known error term distribution of the discrete choice model for each $n \in N$ and $i \in I$, corresponding to different behavioral scenarios. For each scenario $r \in R$, the drawn error term parameter ξ_{inr} is included in the utility function as follows:

$$U_{inr} = V_{in} + \xi_{inr}, \quad (2)$$

and consumers deterministically choose the alternative with the highest utility. This means that the utility of the chosen alternative is equal to

$$U_{nr}^{max} = \max_{j \in I} U_{jnr}. \quad (3)$$

Then, we can express the deterministic choice of consumer $n \in N$ in a specific scenario $r \in R$ using the binary variable x_{inr} as follows:

$$x_{inr} = \begin{cases} 1 & \text{if } U_{inr} = U_{nr}^{max}, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

Furthermore, we consider a supplier k that controls a set of alternatives $I_k \subset I$. The supplier has control over some attributes of these alternatives in order to optimize its objective function, which in this discussion we assume to be related to profit maximization. In Section 2.1 we look at a pricing problem, while in Section 2.2 we analyze a combined assortment and pricing problem.

2.1 Pricing

Initially, we consider the case in which the supplier only controls the prices p_i at which its alternatives $i \in I_k$ are offered. For the sake of this discussion, no price differentiation is applied across customers. In this case, the deterministic part of the utility functions can be expressed as $U_{inr} = \beta_{p,inr} p_i + \hat{q}_{inr}$, where the term $\hat{q}_{inr}$ gathers all the socioeconomic characteristics and all the attributes of the alternative which are not directly affected by the decisions of the supplier.

2.1.1 Continuous price variables

Prices are modeled as lower and upper bounded continuous variables. Throughout this work, we assume that all utilities U_{inr} are positive for each $i \in I$, $n \in N$ and $r \in R$. This can be obtained with a translation of the utility functions U_{inr} that is equal across alternatives for each $n \in N$ and $r \in R$. The translation is always possible when all variables are bounded.

Then, the *Continuous Pricing Problem* (CPP) of the supplier can be written as follows:

$$\max_{p} \quad \pi = \sum_{i \in I_k} \sum_{n \in N} \sum_{r \in R} \frac{1}{|R|} \theta_n p_i x_{inr}, \tag{5}$$

$$s.t. \quad \sum_{i \in I} x_{inr} = 1 \qquad \forall n \in N, \forall r \in R, \tag{6}$$

$$U_{inr} = \beta_{p,inr} p_i + q_{inr} + \xi_{inr} \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{7}$$

$$\sum_{j \in I} U_{jnr} x_{jnr} \geq U_{inr} \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{8}$$

$$0 \leq p_i \leq M_i^p \qquad \forall i \in I, \tag{9}$$

$$x_{inr} \in \{0, 1\} \qquad \forall i \in I, \forall n \in N, \forall r \in R. \tag{10}$$

The objective function (5) maximizes profits. For the sake of simplicity, here we neglect all fixed and variable costs which the supplier might incur. Constraints (7) defines the price-dependent utility functions for each alternative, customer and scenario. Constraints (6) impose that in each scenario every customer chooses one alternative. Notice that choice is modeled through the binary constraints x_{inr}. Constraints (8) state that the chosen alternative must be the one maximizing utility.

We can see that the objective function (5) and the set of constraints (8) are non-linear, as they include a product of a binary and a continuous variable. Let us define the auxiliary continuous variable $w_{inr} = p_i \cdot x_{inr}$, which allows to linearize the product in the following manner:

$$
\begin{aligned}
0 \leq w_{inr} \leq p_i &\qquad \forall i \in I, \forall n \in N, \forall r \in R, &(11)\\
w_{inr} \leq M_i^p x_{inr} &\qquad \forall i \in I, \forall n \in N, \forall r \in R, &(12)\\
p_i - (1 - x_{inr}) M_i^p \leq w_{inr} &\qquad \forall i \in I, \forall n \in N, \forall r \in R. &(13)
\end{aligned}
$$

Then, model (5)-(10) can be written as a mixed integer linear optimization model as follows:

$$
\begin{aligned}
\max_p \quad & \pi = \sum_{i \in I_k} \sum_{n \in N} \sum_{r \in R} \frac{1}{|R|} \theta_n \alpha_{inr}, && (14)\\
s.t. \quad & \sum_{i \in I} x_{inr} = 1 & \forall n \in N, \forall r \in R, \quad & (15)\\
& U_{inr} = \beta_{p,inr} p_i + q_{inr} + \xi_{inr} & \forall i \in I, \forall n \in N, \forall r \in R, \quad & (16)\\
& U_{inr} \leq U_{nr} & \forall i \in I, \forall n \in N, \forall r \in R, \quad & (17)\\
& U_{nr} \leq U_{inr} + M_{nr}^U (1 - x_{inr}) & \forall i \in I, \forall n \in N, \forall r \in R, \quad & (18)\\
& p_i - (1 - x_{inr}) M_i^p \leq w_{inr} \leq p_i & \forall i \in I, \forall n \in N, \forall r \in R, \quad & (19)\\
& w_{inr} \leq M_i^p x_{inr} & \forall i \in I, \forall n \in N, \forall r \in R, \quad & (20)\\
& p_i \leq M_i^p & \forall i \in I, \quad & (21)\\
& p_i \geq 0 & \forall i \in I, \quad & (22)\\
& w_{inr} \geq 0 & \forall i \in I, \forall n \in N, \forall r \in R. \quad & (23)\\
& x_{inr} \in \{0, 1\} & \forall i \in I, \forall n \in N, \forall r \in R, \quad & (24)
\end{aligned}
$$

Notice that constraints (17)-(18) are a linear reformulation of the utility maximization constraints (8). Although not desirable from a computational perspective, the use of big-M constraints is necessary to linearize the formulation. Model (14-24) can be solved as such, for instance by using a general MILP solver.

2.1.2 Discrete price variables

To circumvent the issue of non-linearity, which requires the use of big-M constraints, we explore here a different approach to model the interdependence between consumer utilities and supplier profits. The main additional assumption is that all the decision variables of the supplier can only take a finite set of values. In the case of continuous variables such as prices, it is therefore necessary to identify a meaningful discretization, whose consequences must be evaluated in light of the problem to be solved.

Starting from the non-linear model (5)-(10), for each alternative $i \in I_k$ we constrain prices p_i to belong to the set $Q_i = \{p_i^1, p_i^2, ..., p_i^{|Q|}\}$. This can be done by expanding the set of alternatives I and creating from each original alternative i one alternative for every price level $p \in Q_i$. We define the expanded set of each alternative as I_i^{exp} and the universal expanded choice set as $I^{exp} = \bigcup_i I_i^{exp} \cup (I \setminus I_k)$. All the utility functions defined for each customer $n \in N$ and alternative $i \in I^{exp}$ are now parameters of the optimization model, since they can be expressed as

$$\hat{U}_{inr} = \beta_{p,inr}\hat{p}_i + \hat{q}_{inr} + \xi_{inr} \quad \forall i \in I^{exp}, \forall n \in N, \forall r \in R.$$

Therefore, the *Discrete Pricing Problem* (DPP) of the supplier can be written as follows:

$$\max_{y} \quad \pi = \sum_{i \in I_k} \sum_{n \in N} \sum_{r \in R} \frac{1}{|R|} \theta_n \hat{p}_i x_{inr}, \tag{25}$$

$$s.t. \quad \sum_{j \in I_i^{exp}} y_j = 1 \qquad \forall i \in I, \tag{26}$$

$$\sum_{i \in I^{exp}} x_{inr} = 1 \qquad \forall n \in N, \forall r \in R, \tag{27}$$

$$x_{inr} \leq y_i \qquad \forall i \in I^{exp}, \forall n \in N, \forall r \in R, \tag{28}$$

$$\sum_{j \in I^{exp}} \hat{U}_{jnr} x_{jnr} \geq \hat{U}_{inr} y_i \qquad \forall i \in I^{exp}, \forall n \in N, \forall r \in R, \tag{29}$$

$$x_{inr} \in \{0, 1\} \qquad \forall i \in I^{exp}, \forall n \in N, \forall r \in R, \tag{30}$$

$$y_i \in \{0, 1\} \qquad \forall i \in I^{exp}. \tag{31}$$

The objective function (25) maximizes profits. Constraints (26) require the supplier to choose one price level for each alternative. This is enforced through the set of binary variables y_i. Constraints (27)-(28) ensure that in each scenario every customer chooses one alternative, and the chosen alternative must correspond to a price level chosen by the supplier. Constraints (29) impose that the chosen alternative must be the available alternative that maximizes utility.

Furthermore, we can express the lower-level utility maximization problem for a single customer n and scenario r as follows:

$$\max_x \quad U = \sum_{i \in I} \hat{U}_i x_i, \tag{32}$$

$$s.t. \quad \sum_{i \in I} x_i = 1, \tag{33}$$

$$x_i \leq y_i^* \qquad \forall i \in I, \tag{34}$$

$$x_i \geq 0 \qquad \forall i \in I, \tag{35}$$

where the indexes n and r have been dropped for the sake of simplicity. We notice that the constraint matrix of problem (32)-(35) is totally unimodular. This convenient property allows to relax the integrality constraints (30) on the choice variables x_{inr}, which can simply be replaced by the non-negativity constraints

$$x_{inr} \geq 0 \qquad \forall i \in I^{exp}, \forall n \in N, \forall r \in R. \tag{36}$$

This means that the customer's utility maximization problem for each customer n and scenario r is a continuous knapsack problem where the knapsack's capacity is equal to 1 and each item (alternative) i has a weight of 1 and a profit of $\hat{U}_{inr}$. This property will be exploited in Section 4 when deriving a Benders decomposition scheme for the choice-based optimization with discrete variables.

2.2 Assortment and pricing

Another aspect worth evaluating when discussing the differences between choice-based optimization models with continuous and discrete price variables is the possibility to incorporate other decision variables in the formulation, and the impact that these decisions have from a computational perspective. Here, we consider the case of assortment, that is, the decision about whether or not to offer any given product to the customers. In many optimization problems, this is a strategic decision which is made before the pricing stage, in a sequential manner. However, in other application it can be convenient to treat assortment (or, equivalently, in Section 2.3.2, facility location) and pricing as simultaneous decisions.

2.2.1 Assortment and continuous price variables

To also consider assortment decisions, model (14)-(24) must be modified to include a new set of auxiliary variables $U^a_{inr} = U_{inr} \cdot y_i$, which are needed to model the fact that the customer must choose the alternative with the highest utility among those that are made available by the supplier. This yields the following formulation for the *Assortment and Continuous Pricing Problem* (ACPP):

$$\max_{p} \quad \pi = \sum_{i \in I_k} \sum_{n \in N} \sum_{r \in R} \frac{\theta_n}{|R|} w_{inr}, \tag{37}$$

$$\text{s.t.} \quad \sum_{i \in I} x_{inr} = 1 \qquad \forall n \in N, \forall r \in R, \tag{38}$$

$$U_{inr} = \beta_{p,inr} p_i + q_{inr} + \xi_{inr} \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{39}$$

$$U^a_{inr} \leq U_{inr} \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{40}$$

$$U_{inr} \leq U^a_{inr} + M_{U_{inr}}(1 - y_i) \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{41}$$

$$U^a_{inr} \leq M^U_{inr} y_i \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{42}$$

$$U^a_{inr} \leq U_{nr} \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{43}$$

$$U_{nr} \leq U^a_{inr} + M_{U_{nr}}(1 - x_{inr}) \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{44}$$

$$p_i \leq M^p_i \qquad \forall i \in I, \tag{45}$$

$$p_i - (1 - x_{inr}) M^p_i \leq w_{inr} \leq p_i \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{46}$$

$$w_{inr} \leq M^p_i x_{inr} \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{47}$$

$$p_i \geq 0 \qquad \forall i \in I, \tag{48}$$

$$x_{inr} \in \{0, 1\} \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{49}$$

$$w_{inr} \geq 0 \qquad \forall i \in I, \forall n \in N, \forall r \in R. \tag{50}$$

2.2.2 Assortment and discretized price variables

To include assortment decisions in model (25)-(31), it is sufficient to remove the set of constraints (26). Indeed, for an alternative $i \in I$, not choosing any of the price levels that define the expanded set I^{exp}_i is equivalent to not including alternative i in the assortment. Therefore, the *Assortment and Discrete Pricing Problem* (ADPP) has (25) as objective function and (27)-(31) as constraints.

Alternative	0	1	2	3	4	5	6	7
Mode	Car	IC	Air	Air	HSR	HSR	HSR	HSR
Endogenous	No	No	No	No	Yes	Yes	Yes	Yes
Dep time	-	2:00	7:10	8:10	5:45	6:45	5:40	6:40
Arr time	-	10:00	8:20	9:20	8:45	9:45	9:00	10:00
Travel time (min)	360	480	70	70	180	180	200	200
Waiting time (min)	-	-	60	60	-	-	-	-
Access time (min)	-	0-60	30-60	30-60	0-60	0-60	0-60	0-60
Egress time (min)	-	0-30	30-60	30-60	0-30	0-30	0-30	0-30
Price (€)	100	30	60	60	p_4	p_5	p_6	p_7

Table 1: Attributes of all scheduled services for the high-speed rail pricing problem instance.

2.3 Computational analysis

We perform some computational experiments to compare the models detailed in Section 2.1 and Section 2.2. We cover two cases: the first case considers prices to be the only decision variables on the supply side (CPP and DPP); the second case combines assortment and pricing (ACPP and ADPP). All MILP models are solved using CPLEX 20.1 with a time limit of 36 hours.

2.3.1 Pricing

Instance description. As a case study, we consider an intercity transport market in which various modes are available to travel between two cities in a typical morning period. In this setting, we take the perspective of a high-speed rail operator wanting to optimize prices in order to maximize its profits. Departure times and travel times of all scheduled alternatives are assumed to be exogenously given, together with the prices of the non-high-speed rail and airline alternatives. Additionally, we include a private transport option. Table 1 presents the attributes of all alternatives in the choice set.

We consider a synthetic population of 1000 travelers, categorized into 12 groups of consumers, each having homogeneous socioeconomic characteristics. Homogeneous groups differ with respect to trip purpose (business or other), income level (high or low) and origin location (urban or rural) which leads to different access times to terminals. Furthermore, each individual has a desired arrival time at destination between 9:00 and 11:00 which follows a non-uniform distribution. The following demand patterns are to be mentioned: most business travelers desire to arrive at their final destination before 10:00, while most other travelers are indifferent to arrival time; there is a higher proportion of high

β	Business travelers			Other purpose travelers	
μ_{HSR}	1.190			1.333	
μ_{Air}	1.086			1.106	
ASC_{Car}	0.000*			0.000*	
ASC_{IC}	-1.289*			-2.138*	
ASC_{Air}	-2.893*			-1.856*	
ASC_{HSR}	-0.825*			-0.572*	
Travel time (min)	-0.0133			-0.0054	
Access/egress time (min)	-0.00555			-0.0103	
Early schedule delay (min)	-0.00188			-0.00677	
Late schedule delay (min)	-0.0130			-0.00617	
	Reimbursed	High income	Low income	High income	Low income
Cost car (euro)	-0.0222*	-0.0296*	-0.0527	-0.0228*	-0.0405
Cost Air (euro)	-0.0109	-0.0113*	-0.0201	-0.0109*	-0.0194
Cost IC (euro)	-0.0158	-0.0212*	-0.0377	-0.0097*	-0.0172
Cost HSR (euro)	-0.0120	-0.0160*	-0.0284	-0.0144*	-0.0256

Table 2: Discrete choice model parameters used in the numerical experiments.

income and business travelers among urban travelers than among rural travelers; a part of business travelers are reimbursed and are therefore less price sensitive. For the discrete choice model, we refer to the model estimated by Cascetta and Coppola (2012) from an intercity travel survey conducted in Italy. Table 2 illustrates the parameters used in our experiments. Two separate sets of parameters are considered for business trips and other trip purposes. Additionally, the cost parameters are mode-specific and interact with income, producing different values of travel time savings. A nested logit model is used where two nests μ_{HSR} and μ_{Air} capture the correlation between the scheduled services of the train operator and of the airline. A more detailed description of the input data can be found in Bortolomiol *et al.* (2021b). We remark that the dataset used for the experiments and the derived results are hypothetical and do not represent real scenarios that are related to choices made by existing high-speed rail operators.

Numerical results. Here, we are interested in solving the optimization problem of the high-speed rail operator. The decision variables are the prices p_4, p_5, p_6 and p_7 of the four high-speed rail departures. We execute experiments on models (14)-(24) and (25)-(31) by varying the following parameters: (i) the number of simulation scenarios $|R| = 20, 50, 100, 200$; (ii) the lower bound on the price variables $p_i \geq 0$ € or $p_i \geq 100$ €; (iii) for the model with discretized price levels, the size of the set Q_i of prices, with $|Q_i| = 21, 51, 101$, assumed to be equal for each alternatives $i \in I_k$, which cover the feasible range of the continuous problem by defining evenly spaced values between lower bound and upper bound (which is set to 200 € for all alternatives and instances).

Table 3 shows the results of these numerical tests. As expected, computational times increase for both the CPP and the DPP when the number of scenarios increases and decrease when the bounds on the price variables are tighter. The latter observation highlights the importance of providing tight variables bounds by excluding irrelevant regions of the search space in order to increase the speed of convergence. The running time of the DPP increases exponentially with the size of the expanded sets I_i^{exp}. A comparison between the CPP and the DPP shows that the DPP converges to optimality faster than the CCP when $|R| \geq 100$ and $|I_i^{exp}| = 21$. It is also important to notice that the CPP generalizes the DPP for any size of the discretized sets I_i^{exp}, therefore the optimal solution of the CPP is an upper bound (in a profit maximization context) of the optimal solution of the DPP. The gap between the two optimal solutions depends on the chosen discretization. Here, it is never higher than 0.55% when $UB - LB = 100$ € and $|I_i^{exp}| = 101$, that is, when we allow the discretized price variables to take any integer value between 100 € and 200 €.

Instance			CPP						DPP							Gap
$\lvert R\rvert$	LB	UB	Time	Opt	p_4	p_5	p_6	p_7	$\lvert I_i^{exp}\rvert$	Time	Opt	p_4	p_5	p_6	p_7	
									21	6.75	70072.00	130.00	100.00	140.00	150.00	2.37%
20	0	200	4.35	71774.95	134.19	100.67	144.95	156.47	51	24.07	70885.60	132.00	100.00	144.00	152.00	1.24%
									101	61.01	71316.20	134.00	100.00	144.00	154.00	0.64%
									21	1.42	70390.50	130.00	100.00	140.00	155.00	1.93%
20	100	200	0.45	71774.95	134.19	100.67	144.95	156.47	51	7.18	71316.20	134.00	100.00	144.00	154.00	0.64%
									101	8.89	71379.90	134.00	100.00	144.00	155.00	0.55%
									21	53.62	71720.00	200.00	150.00	110.00	160.00	0.97%
50	0	200	55.01	72423.71	189.67	152.25	110.58	161.24	51	246.11	71589.60	200.00	152.00	112.00	160.00	1.15%
									101	1262.30	72106.36	200.00	152.00	110.00	160.00	0.44%
									21	14.59	71889.00	200.00	135.00	110.00	160.00	0.74%
50	100	200	10.46	72423.71	189.67	152.25	110.58	161.24	51	31.51	72106.36	200.00	152.00	110.00	160.00	0.44%
									101	89.91	72185.30	189.00	152.00	110.00	161.00	0.33%
									21	163.57	65874.50	130.00	180.00	160.00	120.00	1.27%
100	0	200	461.74	66724.56	130.09	188.50	139.31	89.88	51	1026.32	66098.60	124.00	176.00	148.00	120.00	0.94%
									101	5578.19	66255.90	130.00	188.00	148.00	120.00	0.70%
									21	34.48	66118.40	125.00	175.00	155.00	120.00	0.50%
100	100	200	101.64	66452.18	130.09	188.50	139.31	121.38	51	161.03	66255.90	130.00	188.00	148.00	120.00	0.30%
									101	395.86	66341.32	130.00	188.00	139.00	121.00	0.17%
									21	717.89	69543.55	140.00	110.00	130.00	120.00	1.76%
200	0	200	1824.03	70788.17	135.73	108.29	139.73	108.32	51	3746.48	70343.40	136.00	108.00	132.00	108.00	0.63%
									101	46337.20	70489.95	126.00	108.00	138.00	108.00	0.42%
									21	139.17	69859.60	125.00	115.00	135.00	110.00	1.31%
200	100	200	288.89	70788.17	135.73	108.29	139.73	108.32	51	415.90	70489.95	126.00	108.00	138.00	108.00	0.42%
									101	1829.24	70571.67	126.00	107.00	139.00	108.00	0.31%

Table 3: Results for the high-speed rail case study when using the CPP and the DPP to solve the supplier's pricing problem to optimality.

β	Value
ASC_{FSP}	0.0
ASC_{PSP}	32.0
ASC_{PUP}	34.0
Fee (€)	$\sim \mathcal{N}(-32.328, 14.168)$
Fee PSP - low income (€)	-10.995
Fee PUP - low income (€)	-13.729
Travel time to parking (min)	$\sim \mathcal{N}(-0.788, 1.06)$
Travel time to destination (min)	-0.612
Age of vehicle (1/0)	4.037
Origin (1/0)	-5.762

Table 4: Discrete choice model parameters used in the numerical experiments.

2.3.2 Assortment and pricing

Instance description. For this set of experiments, we look at the combined problem of selecting a number of sites where to open parking facilities among a set of candidate locations and determining a price that customers must pay to access open facilities, with the goal of maximizing the profits of the operator.

We consider a graph representing a stylized city road network, presented in Figure 1. The supplier considers 8 locations where facilities can be opened, corresponding to the blue diamond in the graph. The demand is constituted of commuters who want to travel from the peripheral regions to the city center. In the graph, the origins of the commuters are represented with red circles, while the center is represented with a yellow star. Commuters have heterogeneous socioeconomic characteristics, which lead to different preferences. We use the discrete choice model estimated by Ibeas *et al.* (2014), who use a mixed logit model to study car driver's behavior when choosing among three different parking alternatives available in a small Spanish town. The explanatory socioeconomic variables include trip origin, age of the vehicle and income level. Additionally, the following attributes of the alternatives are considered: type of parking (underground, indicates with U in the figure, or on-street, indicated with S), travel time from parking to destination, travel time from origin to parking and parking fee. For the latter two continuous variables, the corresponding coefficients are normally distributed in the utility function. Table 4 illustrates the parameters of the discrete choice model derived from Ibeas *et al.* (2014).

Numerical results. The supplier's decision variables include the binary choice of opening or not each of the candidate facilities and the prices of the opened facilities. We perform experiments on models (37)-(50) and (25)+(27)-(31) by varying the number of simulation scenarios $|R| = 10, 20, 50, 100$ and the size of the sets Q_i of discretized prices.

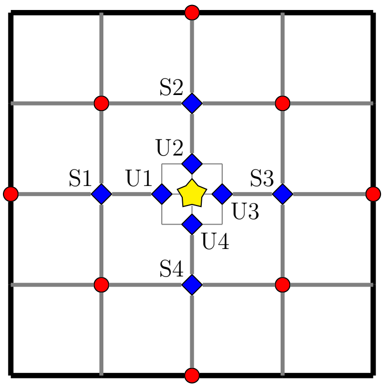

Figure 1: Parking locations used in the case study

Instance			ACPP		ADPP			Gap
$\lvert R\rvert$	LB	UB	Time	Opt	$\lvert I_i^{exp}\rvert$	Time	Opt	
10	0.00	3.00	11706	907.8	16	132	864.0	4.82%
					31	800	876.0	3.50%
20	0.00	3.00	129600*	877.0*	16	429	842.0	3.99%
					31	2778	862.5	1.65%
50	0.00	3.00	129600*	842.8*	16	837	816.4	3.13%
					31	12191	830.4	1.47%
100	0.00	3.00	129600*	844.0*	16	3419	828.2	1.87%
					31	39425	831.8	1.45%

Table 5: Results for the parking case study when using the ACPP and the ADPP to solve the supplier's assortment and pricing problem to optimality.

Table 5 shows the results of these numerical tests. Results marked with an asterisk indicate that optimality was not proven within the time limit of 36 hours. The major finding from this set of experiments is that the presence of the binary assortment variables has a greater impact on the ACPP than on the ADPP. Indeed, even a relatively small instance with $|I_k| = 8$, $|N| = 8$ and $|R| = 20$ cannot be solved to optimality within 36 hours. By observing the progression of the CPLEX log during the execution, we notice that for both models good solutions are found early in the algorithm and that most of the subsequent effort goes into closing the optimality gap.

2.3.3 Discussion

These numerical experiments show the trade-offs to be considered when choosing between choice-based optimization models that do or do not include continuous variables at the supply level. In particular, we see that a complex decision space which includes both continuous and discrete variables (price and assortment, in our case) cannot be handled efficiently by the MILP solver. In this case, a model where all upper-level variables are discrete provides a valuable alternative for two reasons: (i) the discretization of the continuous variables can be informed by problem-specific heuristics, and the algorithm can be designed to control the level of approximation in the discretization; (ii) the resulting formulation, in which the utility functions of the expanded choice set are parameters, is such that the lower-level optimization problem of the customer is a continuous knapsack problem. The latter observation is particularly relevant for the development of a Benders decomposition algorithm, which is the focus of the following sections.

3 Benders decomposition

3.1 Classical Benders decomposition

While Benders decomposition can also be applied to more generic optimization problems, here we focus on mixed integer linear programming (MILP) to present the classical version of the Benders decomposition algorithm (Benders, 1962). Consider the following optimization problem:

$$\min_{y,x} \quad f^T y + c^T x, \tag{51}$$

$$s.t. \quad Ay = b, \tag{52}$$

$$By + Dx = d, \tag{53}$$

$$x \geq 0, \tag{54}$$

$$y \in \mathbb{Z}^+, \tag{55}$$

where $x \in \mathbb{R}^{n_2}$, $y \in \mathbb{Z}^{n_1}$, $A \in \mathbb{R}^{m_1 \times n_1}$, $B \in \mathbb{R}^{m_2 \times n_1}$, $D \in \mathbb{R}^{m_2 \times n_2}$, $b \in \mathbb{R}^{m_1}$, $d \in \mathbb{R}^{m_2}$.

Model (51)-(55) can be rewritten as follows:

$$\min_{\bar{y} \in \mathbb{Z}^{n_1}} \left\{ f^T \bar{y} + \min_{x \geq 0} \{c^T x : Dx = d - B\bar{y}\} \right\}. \tag{56}$$

In (56), $\bar{y}$ is a feasible solution for the complicating integer variables. When we assume that $\bar{y}$ is fixed, the inner minimization problem is a continuous problem that can be dualized as follows:

$$\max_{\alpha \in \mathbb{R}^{m_2}} \left\{ (d - B\bar{y})^T \alpha : D\alpha \geq c \right\}, \tag{57}$$

where the variables α are associated to the primal constraints $Dx = d - B\bar{y}$. From strong duality, we know that formulation (56) is equivalent to the following formulation:

$$\min_{\bar{y} \in \mathbb{Z}^{n_1}} \left\{ f^T \bar{y} + \max_{\alpha} \{ (d - B\bar{y})^T \alpha : D\alpha \geq c \} \right\}. \tag{58}$$

Notice that in (58) the feasible region F of the inner maximization problem is independent from $\bar{y}$. If the primal problem is feasible, then (57) can be feasible or unbounded. Here, we restrict our discussion to the former case, in which the solution belongs to the set E of the extreme points of F. By assuming dual feasibility, we can reformulate (58) using an artificial continuous variable as follows:

$$\min_{y,z} \quad f^T y + z, \tag{59}$$

$$s.t. \quad Ay = b, \tag{60}$$

$$z \geq (d - B\bar{y})^T \alpha_e, \qquad \forall e \in E \tag{61}$$

$$y \in \mathbb{Z}^+, \tag{62}$$

This formulation is commonly referred to as Benders *master problem* (MP). The set of constraints (61) are known as Benders *optimality cuts* and determine a lower bound of the contribution to the objective function of the original continuous variables x, which have been projected away. To avoid full enumeration, an iterative approach is proposed in which the optimality cuts are initially excluded from the model and then progressively added to the *restricted master problem* (RMP) using a dynamic cutting-plane generation technique that consists in solving the RMP to obtain a trial value of the integer variables $\bar{y}$ and then solving the *worker problem* (57) with $\bar{y}$ to get the dual variables α_e and produce a valid cut to add to the set (61). This approach constitutes the classical implementation of Benders decomposition.

3.2 Issues and enhancements

Although Benders decomposition is guaranteed to converge to an optimal solution in a finite number of iterations, the classical implementation is known to be inefficient for a number of reasons. Here, we outline the most relevant issues and we discuss the corresponding enhancements that have been proposed in the literature. A more extensive treatment of the algorithmic properties of Benders decomposition and its variations can be found in the review paper by Rahmaniani *et al.* (2017).

Master problem In the classical Benders decomposition, a restricted master problem, whose size increases after each iteration, is solved at each iteration to obtain the optimal integer variables that is used in the dual subproblem. For this reason, the master problem usually represents the bottleneck of the algorithm. Some alternative strategies that have been proposed in the literature to increase computational speed are the following: not solving the RMP to optimality at each iteration, since in principle a feasible solution is sufficient to generate Benders cuts at the subproblem level (Geoffrion and Graves, 1974); using a cut selection strategy that only adds those cuts that improve the best-known upper bound (Rei *et al.,* 2009); adding cuts as *lazy constraints*, that is, keeping them in a pool of constraints which are added only when they are violated by an incumbent solution. A more modern and advanced approach, which has been shown to consistently outperform the classical Benders decomposition, avoids solving a new MILP at each iteration by incorporating the Benders decomposition into a general branch-and-cut algorithm (Fortz and Poss, 2009, Ljubić *et al.,* 2012, Fischetti *et al.,* 2016). In this way, a single branch-and-bound enumeration tree is generated for the initial restricted master problem, and Benders cuts are separated on the fly while processing the nodes of the tree. This method is typically referred to as branch-and-Benders-cut algorithm. An advantage of this method is the possibility to integrate Benders cuts and other cutting planes, by exploiting problem-specific information while processing the branch-and-bound tree. Furthermore, the branch-and-Benders-cut implementation offers potential speed-up opportunities in terms of heuristic strategies related to branching rules, node selection and pruning.

Worker problem The worker problem is a linear problem that can be solved to optimality using the simplex or other well-known algorithms. Two aspects are worth discussing here. The first aspect is related to the existence of a block-diagonal structure that allows decomposing the worker problem into smaller independent subproblems. This possibility often justifies the use of the Benders decomposition in the first place, and is particularly relevant for stochastic optimization problems, which will be treated in depth in Section 3.3. In this case, each subproblem $r \in R$ will provide a set of dual variables α_r that can be used to generate the following outer linearization approximation to be added to the restricted master problem:

$$z^r \geq (d^r - B^r \bar{y})^T \alpha_e^r. \tag{63}$$

Notice that the resulting disaggregate cuts can either be added as such or else be fully or partially aggregated (Birge and Louveaux, 1988). The efficiency of different cut bundling strategies within a multi-cut scheme is problem-dependent. On the one hand, disaggregate cuts provide subproblem-specific information that can help cutting the solution space more efficiently. On the other hand, adding too many cuts, especially if redundant, might slow down the algorithm. An example of computational analysis for a public transport network design problem can be found in Mahéo *et al.* (2019), where intermediate bundling approaches based on grouping strategies that are informed by problem inputs are found to outperform both the fully disaggregated approach and the single-cut approach. The second aspect is related to the selection of the solution of the worker problem from which Benders cut are generated. This is especially relevant when the worker problem is degenerate and, therefore, different cuts could be obtained from different optimal solutions. Magnanti and Wong (1981) introduced the concept of Pareto-optimal cuts to identify those cuts that are not dominated for any feasible $\bar{y}$ and showed that such cuts can be derived by solving an auxiliary problem that uses a core point y_0 of the set Y, that is, a point in the relative interior of the convex hull defined by the feasible points of Y. This approach has been further investigated by Papadakos (2008) and Sherali and Lunday (2013), among others, who proposed enhancements to the Magnanti-Wong method.

3.3 Benders decomposition and stochastic optimization

Stochastic optimization methods rely upon a finite set of representative scenarios to approximate the possible outcomes for the values of the stochastic parameters (Crainic *et al.*, 2021). In transport optimization problems, uncertainty can exist on parameters such as expected demand, travel time and travel cost. In this context, solutions are evaluated under each scenario and weighted according to the probability of occurrence of the scenario, and their overall quality is then a result of some form of aggregation. The need to generate large number of scenarios to represent uncertainty produces large-scale models which are characterized by sets of variables that are duplicated in each independent scenario. These variables are referred to as second-stage variables, as opposed to scenario-independent variables which are known as first-stage variables. The results is a block-diagonal structure, which makes Benders decomposition a promising solution approach when second-stage variables are linear. Indeed, this technique, also called L-shaped method, has been used for decades in stochastic programming (see Van Slyke and Wets (1969), Laporte and Louveaux (1993), Birge and Louveaux (2011)). From a computational perspective, recent efforts to improve classical Benders decomposition for stochastic optimization focus on exploiting information from relevant scenarios. This is motivated by the observation that, when all second-stage information is removed, the initial master problem is weak, leading to computational instability and slow convergence until a sufficient number of cuts is added. Crainic *et al.* (2021) introduce partial Benders decomposition, a methodology that aims at including some information from the scenario subproblems in the master problem. The authors propose various data-driven scenario retention and scenario creation strategies which are then tested on a stochastic network design problem. Their results show that a combination of these techniques allows to find better bounds, reduce the size of the explored branch-and-bound trees and the number the Benders cuts needed to prove optimality. Somehow related is the work by Hewitt *et al.* (2021), who propose to identify structures in the scenario space by means of clustering methods. More specifically, once a set R of scenarios is defined, opportunity costs are computed which measure the loss encountered by taking the decision associated to, say, scenario r_1 when scenario r_2 actually occurs. Then, using an opportunity cost distance function, scenarios can be compared and clustered on a decisional basis in order to derive valid upper and lower bounds from a reduced number of scenarios and solutions.

3.4 Applications of Benders decomposition

Benders decomposition has proven to be a go-to methodology for several optimization problems, which can be classified into three main non-mutually exclusive categories: (i) planning problems where strategic decision variables, e.g. location or routing, are integer and operational decision variables, e.g. prices or quantities, are continuous; (ii) stochastic problems which require the evaluation of multiple scenarios which are only connected at the level of the first-stage variables; (iii) bilevel problems where upper-level variables capture the decisions of the leader and lower-level variables those of the follower. A non-exhaustive list of applications of Benders decomposition includes deterministic and stochastic facility location (Tang *et al.,* 2013*,* Fischetti *et al.,* 2017*,* Lin and Tian*,* 2021*,* Parragh *et al.,* 2021), hub location (Contreras *et al.,* 2011), production routing under uncertainty (Adulyasak *et al.,* 2015), network design problems (Binato *et al.,* 2001*,* Costa*,* 2005*,* Fortz and Poss*,* 2009*,* Fontaine and Minner*,* 2014*,* 2018*,* Mahéo *et al.,* 2019*,* Crainic *et al.,* 2021), charging station location problem (Arslan and Karaşan*,* 2016), electric location-routing problem (Çalık *et al.,* 2021).

4 Benders decomposition for the choice-based optimization problem with discrete variables

In this section, we derive a Benders decomposition scheme for the ADPP introduced in Section 2.2.2. This derivation can be easily generalized to any choice-based optimization problem where simulation is used to approximate the choice probabilities of the customers and where all the decision variables of the supplier are discrete.

Let us start from the MILP model (25)+(27)-(31), which describes the bilevel optimization problem of the supplier, and the LP model (32)-(35), which describes the lower-level optimization problem of a single customer. First, we rewrite the latter problem as a minimization problem by changing the sign of the objective function (32):

$$\min_{x} \quad \sum_{i\in I^{exp}} -\hat{U}_i x_i. \tag{64}$$

Dual customer subproblem. Then, the dual of the customer optimization problem is derived by defining the variable α^1, corresponding to constraint (33) of the primal, and the set of variables α_i^2 for each $i \in I^{exp}$, corresponding to the set of constraints (34) of the primal. The dual problem looks as follows:

$$\begin{aligned}
\max_{\alpha^1,\alpha^2} \quad & \alpha^1 + \sum_{i\in I^{exp}} y_i^* \alpha_i^2, && && (65)\\
s.t. \quad & \alpha^1 + \alpha_i^2 \leq -\hat{U}_i && \forall i \in I^{exp}, && (66)\\
& \alpha^1 \leq 0, && && (67)\\
& \alpha_i^2 \leq 0 && \forall i \in I^{exp}. && (68)
\end{aligned}$$

Strong duality conditions state that the primal optimal objective and the dual optimal objective are equal, that is,

$$\sum_{i\in I^{exp}} -\hat{U}_i x_i = \alpha^1 + \sum_{i\in I^{exp}} y_i^* \alpha_i^2. \tag{69}$$

Single-level problem. Thanks to duality, we can then rewrite the utility maximization conditions (29) and obtain the following formulation which is equivalent to model (25)-(31):

$$
\begin{aligned}
\max_{y} \quad & \pi = \sum_{i \in I_k^{exp}} \sum_{n \in N} \sum_{r \in R} \theta_n \hat{p}_i x_{inr}, && && (70)\\
s.t. \quad & \sum_{j \in I_i^{exp}} y_j = 1 && \forall i \in I, && (71)\\
& \sum_{i \in I^{exp}} x_{inr} = 1 && \forall n \in N, \forall r \in R, && (72)\\
& x_{inr} \leq y_i && \forall i \in I^{exp}, \forall n \in N, \forall r \in R, && (73)\\
& \sum_{i \in I^{exp}} -\hat{U}_{inr} x_{inr} = \alpha^1_{nr} + \sum_{i \in I} y_i \alpha^2_{inr} && \forall i \in I^{exp}, \forall n \in N, \forall r \in R, && (74)\\
& -\alpha^1_{nr} - \alpha^2_{inr} \leq -\hat{U}_{inr} && \forall i \in I^{exp}, \forall n \in N, \forall r \in R, && (75)\\
& \alpha^1_{nr} \geq 0 && \forall n \in N, \forall r \in R, && (76)\\
& \alpha^2_{inr} \geq 0 && \forall i \in I^{exp}, \forall n \in N, \forall r \in R, && (77)\\
& x_{inr} \geq 0 && \forall i \in I^{exp}, \forall n \in N, && (78)\\
& y_i \in \{0, 1\} && \forall i \in I_k^{exp}. && (79)
\end{aligned}
$$

Linearized single-level problem. The product $y_i \cdot \alpha^2_{inr}$ in constraints (74) can be linearized as in (11)-(13). We use the auxiliary variables $\delta_{inr} = y_i \cdot \alpha^2_{inr}$ and write the following set of linear constraints:

$$
\begin{aligned}
\delta_{inr} &\leq \alpha^2_{inr} && \forall i \in I^{exp}, \forall n \in N, \forall r \in R, && (80)\\
\delta_{inr} &\geq \alpha^2_{inr} - M_{inr}(1 - y_i) && \forall i \in I^{exp}, \forall n \in N, \forall r \in R, && (81)\\
\delta_{inr} &\leq M_{inr} y_i && \forall i \in I^{exp}, \forall n \in N, \forall r \in R. && (82)
\end{aligned}
$$

And we obtain the following mixed integer linear optimization model:

$$
\begin{aligned}
\max \quad & \pi = \sum_{i \in I_k} \sum_{n \in N} \sum_{r \in R} \frac{1}{|R|} \theta_n \hat{p}_i x_{inr}, && && (83)\\
s.t. \quad & \sum_{j \in I_i^{exp}} y_j = 1 && \forall i \in I, && (84)\\
& \sum_{i \in I^{exp}} x_{inr} = 1 && \forall n \in N, \forall r \in R, && (85)\\
& x_{inr} \leq y_i && \forall i \in I, \forall n \in N, \forall r \in R, && (86)\\
& \sum_{i \in I^{exp}} -\hat{U}_{inr} x_{inr} = \alpha^1_{nr} + \sum_{i \in I} \delta_{inr} && \forall i \in I, \forall n \in N, \forall r \in R, && (87)\\
& -\alpha^1_{nr} - \alpha^2_{inr} \leq -\hat{U}_{inr} && \forall i \in I, \forall n \in N, \forall r \in R, && (88)\\
& \delta_{inr} \leq \alpha^2_{inr} && \forall i \in I, \forall n \in N, \forall r \in R, && (89)
\end{aligned}
$$

$$\delta_{inr} \geq \alpha^2_{inr} - M_{inr}(1 - y_i) \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{90}$$

$$\delta_{inr} \leq M_{inr} y_i \qquad \forall i \in I, \forall n \in N, \forall r \in R. \tag{91}$$

$$\alpha^1_{nr} \geq 0 \qquad \forall n \in N, \forall r \in R, \tag{92}$$

$$\alpha^2_{inr} \geq 0 \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{93}$$

$$\delta_{inr} \geq 0 \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{94}$$

$$x_{inr} \geq 0 \qquad \forall i \in I, \forall n \in N, \tag{95}$$

$$y_i \in \{0, 1\} \qquad \forall i \in I_k, \tag{96}$$

A valid value for the M_{inr} parameters used in constraints (90)-(91) is $\hat{U}_{inr}$.

Dual of the linearized single-level problem with fixed supply decisions. Let us now fix the decision variables y_i of the supplier, which determine product assortment and prices, to a value y_i^*. We can then derive the worker problem, which is used to find the optimal choices x_{inr} given y_i^*:

$$\min \quad \pi = -\sum_{i \in I_k} \sum_{n \in N} \sum_{r \in R} \frac{1}{|R|} \theta_n \hat{p}_i x_{inr}, \tag{97}$$

$$s.t. \quad \sum_{i \in I^{exp}} x_{inr} = 1 \qquad \forall n \in N, \forall r \in R, \quad (\gamma^1_{nr}) \tag{98}$$

$$x_{inr} \leq y_i^* \qquad \forall i \in I, \forall n \in N, \forall r \in R, \quad (\gamma^2_{inr}) \tag{99}$$

$$\sum_{i \in I^{exp}} -\hat{U}_{inr} x_{inr} - \alpha^1_{nr} - \sum_{i \in I} \delta_{inr} \leq 0 \qquad \forall n \in N, \forall r \in R, \quad (\gamma^3_{nr}) \tag{100}$$

$$-\alpha^1_{nr} - \alpha^2_{inr} \leq -\hat{U}_{inr} \qquad \forall i \in I, \forall n \in N, \forall r \in R, \quad (\gamma^4_{inr}) \tag{101}$$

$$\delta_{inr} - \alpha^2_{inr} \leq 0 \qquad \forall i \in I, \forall n \in N, \forall r \in R, \quad (\gamma^5_{inr}) \tag{102}$$

$$\alpha^2_{inr} - \delta_{inr} \leq M(1 - y_i^*) \qquad \forall i \in I, \forall n \in N, \forall r \in R, \quad (\gamma^6_{inr}) \tag{103}$$

$$\delta_{inr} \leq M y_i^* \qquad \forall i \in I, \forall n \in N, \forall r \in R, \quad (\gamma^7_{inr}) \tag{104}$$

$$\alpha^1_{nr} \geq 0 \qquad \forall n \in N, \forall r \in R, \tag{105}$$

$$\alpha^2_{inr} \geq 0 \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{106}$$

$$\delta_{inr} \geq 0 \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{107}$$

$$x_{inr} \geq 0 \qquad \forall i \in I, \forall n \in N, \forall r \in R. \tag{108}$$

From the primal formulation (97)-(108), we can derive the following dual worker problem:

$$\max \quad \sum_{n \in N} \sum_{r \in R} \Bigg(\gamma^1_{nr} + \sum_{i \in I} y_i^* \gamma^2_{inr} + \sum_{i \in I} -\hat{U}_{inr} \gamma^4_{inr} +$$

$$\sum_{i\in I} M_{inr}(1-y_i^*)\gamma_{inr}^6 + \sum_{i\in I} M_{inr} y_i^* \gamma_{inr}^7 \Bigg), \tag{109}$$

$$s.t. \quad \gamma_{nr}^1 + \gamma_{inr}^2 - \hat{U}_{inr}\gamma_{nr}^3 \leq \frac{1}{|R|}\theta_n \hat{p}_i \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{110}$$

$$\sum_{i\in I} -\gamma_{inr}^3 - \gamma_{inr}^4 \leq 0 \qquad \forall n \in N, \forall r \in R, \tag{111}$$

$$-\gamma_{inr}^4 - \gamma_{inr}^5 + \gamma_{inr}^6 \leq 0 \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{112}$$

$$\gamma_{inr}^5 - \gamma_{inr}^6 + \gamma_{inr}^7 \leq 0 \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{113}$$

$$\gamma_{nr}^1 \leq 0 \qquad \forall n \in N, \forall r \in R, \tag{114}$$

$$\gamma_{inr}^2 \leq 0 \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{115}$$

$$\gamma_{nr}^3 \leq 0 \qquad \forall n \in N, \forall r \in R, \tag{116}$$

$$\gamma_{inr}^4 \leq 0 \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{117}$$

$$\gamma_{inr}^5 \leq 0 \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{118}$$

$$\gamma_{inr}^6 \leq 0 \qquad \forall i \in I, \forall n \in N, \forall r \in R, \tag{119}$$

$$\gamma_{inr}^7 \leq 0 \qquad \forall i \in I, \forall n \in N, \forall r \in R. \tag{120}$$

Benders decomposition algorithm. Having derived the dual worker problem (109)-(120), we can now outline the classical Benders decomposition algorithm for the choice-based optimization problem with discrete variables:

1. Initialize the upper bound $UB = \infty$ and the lower bound $LB = -\infty$ of the master problem.

2. Initialize the restricted master problem:

$$\min_{y,z} \quad z \tag{121}$$

$$s.t. \quad \textit{Domain constraints on the } y \textit{ variables} \tag{122}$$

$$z \geq LB_z. \tag{123}$$

   The values of the master variables y^* can be initialized set to any solution that satisfies the domain constraints, while LB_z is any valid lower bound of z, such as the optimal objective value of the linear relaxation of (25)+(27)-(31).

3. Solve the dual worker problem (109)-(120) for $y = y^*$. Retrieve the optimal dual variables γ_{nr}^{1*}, γ_{inr}^{2*}, γ_{nr}^{3*}, γ_{inr}^{4*}, γ_{inr}^{5*}, γ_{inr}^{6*}, γ_{inr}^{7*}. Let z^{DWP} be the current optimal objective value of the dual problem, and therefore also a valid solution of the primal. Update $UB = \min\{UB, z^{DWP}\}$. (Notice that, given y^*, z^{DWP} can be obtained

without solving an optimization model by computing the choice of each customer n for each scenarios r, which corresponds to the available alternative with the highest exogenous utility.)

4. Using the optimal dual variables, add the following optimality cut to the master problem:

$$z \geq \sum_{n\in N}\sum_{r\in R}\Bigg(\gamma^{1*}_{nr} + \sum_{i\in I}\gamma^{2*}_{inr}y_i + \sum_{i\in I} -\hat{U}_{inr}\gamma^{4*}_{inr} + \sum_{i\in I}\gamma^{6*}_{inr}M_{inr}(1-y_i) + \sum_{i\in I} M_{inr}\gamma^{7*}_{inr}y_i\Bigg). \tag{124}$$

5. Solve the current restricted master problem. Save the solution y^M, z^M. Let $f(y^M, z^M)$ be the current optimal objective value. Update $LB = f(y^M, z^M)$.

6. If $UB - LB \leq \epsilon$, then stop. Else, update $y^* = y^M$ and go to step 3.

5 Future directions

In this paper, we look at a class of choice-based optimization problems in which all decision variables of the supplier are discrete and finite. Numerical experiments show that such a formulation is computationally more efficient than one which includes both continuous and discrete variables. Discretizing continuous variables leads to trade-offs between approximation in the solution and computational speed which must be evaluated on a case-by-case basis. An interesting property of the proposed formulation with discrete upper-level variables is that the lower-level optimization problem of the customer can be expressed a continuous knapsack problem. This allows for a straightforward development of a Benders decomposition algorithm for the problem. After reviewing the literature on Benders decomposition, with a particular attention on stochastic optimization, we derive the Benders decomposition algorithm for the problem at hand.

We envision the following topics to be of particular interest to progress on this research: (i) develop a branch-and-Benders-cut algorithm for the problem and implement enhancements both at the master problem level and at the subproblem level among those discussed in Sections 3.2 and 3.3 that can speed-up the execution of the algorithm; (ii) conduct further experiments to compare the computational performance of our approach against a black-box MIP solver on an uncapacitated facility location and pricing problem with disaggregate demand; (iii) investigate smart discretization techniques, possibly with online updates of the expanded choice sets considered by the supplier; (iv) explore heuristic approaches to exploit data at the scenario level and at the customer level.

6 References

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

Arslan, O. and O. E. Karaşan (2016) A Benders decomposition approach for the charging station location problem with plug-in hybrid electric vehicles, *Transportation Research Part B: Methodological*, **93**, 670–695.

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

Binato, S., M. V. F. Pereira and S. Granville (2001) A new Benders decomposition approach to solve power transmission network design problems, *IEEE Transactions on Power Systems*, **16** (2) 235–240.

Birge, J. R. and F. Louveaux (2011) *Introduction to stochastic programming*, Springer Science & Business Media.

Birge, J. R. and F. V. Louveaux (1988) A multicut algorithm for two-stage stochastic linear programs, *European Journal of Operational Research*, **34** (3) 384–392.

Bortolomiol, S., V. Lurkin and M. Bierlaire (2021a) Price-based regulation of oligopolistic markets under discrete choice models of demand, *Transportation*, **Accepted for Publication** (x) xx–xx.

Bortolomiol, S., V. Lurkin and M. Bierlaire (2021b) A simulation-based heuristic to find approximate equilibria with disaggregate demand models, *Transportation Science*, **Accepted for Publication** (x) xx–xx.

Çalık, H., A. Oulamara, C. Prodhon and S. Salhi (2021) The electric location-routing problem with heterogeneous fleet: Formulation and Benders decomposition approach, *Computers & Operations Research*, **131**, 105251.

Cascetta, E. and P. Coppola (2012) An elastic demand schedule-based multimodal assignment model for the simulation of high speed rail (HSR) systems, *EURO Journal on Transportation and Logistics*, **1** (1-2) 3–27.

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

Costa, A. M. (2005) A survey on benders decomposition applied to fixed-charge network design problems, *Computers & operations research*, **32** (6) 1429–1450.

Crainic, T. G., M. Hewitt, F. Maggioni and W. Rei (2021) Partial Benders decomposition: general methodology and application to stochastic network design, *Transportation Science*, **55** (2) 414–435.

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

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

Fontaine, P. and S. Minner (2014) Benders decomposition for discrete–continuous linear bilevel problems with application to traffic network design, *Transportation Research Part B: Methodological*, **70**, 163–172.

Fontaine, P. and S. Minner (2018) Benders decomposition for the hazmat transport network design problem, *European Journal of Operational Research*, **267** (3) 996–1002.

Fortz, B. and M. Poss (2009) An improved Benders decomposition applied to a multi-layer network design problem, *Operations research letters*, **37** (5) 359–364.

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

Hewitt, M., J. Ortmann and W. Rei (2021) Decision-based scenario clustering for decision-making under uncertainty, *Annals of Operations Research*, 1–25.

Ibeas, A., L. Dell'Olio, M. Bordagaray and J. d. D. Ortúzar (2014) Modelling parking choices considering user heterogeneity, *Transportation Research Part A: Policy and Practice*, **70**, 41–49.

Laporte, G. and F. V. Louveaux (1993) The integer l-shaped method for stochastic integer programs with complete recourse, *Operations research letters*, **13** (3) 133–142.

Lin, Y. H. and Q. Tian (2021) Branch-and-cut approach based on generalized benders decomposition for facility location with limited choice rule, *European Journal of Operational Research*, **293** (1) 109–119.

Ljubić, I., P. Putz and J.-J. Salazar-González (2012) Exact approaches to the single-source network loading problem, *Networks*, **59** (1) 89–106.

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

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

Manski, C. F. (1977) The structure of random utility models, *Theory and decision*, **8** (3) 229–254.

Pacheco, M., M. Bierlaire, B. Gendron and S. S. Azadeh (2021) Integrating advanced discrete choice models in mixed integer linear optimization, *Transportation Research Part B: Methodological*, **146**, 26–49.

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

Parragh, S. N., F. Tricoire and W. J. Gutjahr (2021) A branch-and-Benders-cut algorithm for a bi-objective stochastic facility location problem, *OR Spectrum*, 1–41.

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

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

Sherali, H. D. and B. J. Lunday (2013) On generating maximal nondominated Benders cuts, *Annals of Operations Research*, **210** (1) 57–72.

Tang, L., W. Jiang and G. K. Saharidis (2013) An improved Benders decomposition algorithm for the logistics facility location problem with capacity expansions, *Annals of operations research*, **210** (1) 165–190.

Van Slyke, R. M. and R. Wets (1969) L-shaped linear programs with applications to optimal control and stochastic programming, *SIAM Journal on Applied Mathematics*, **17** (4) 638–663.