Optimization algorithms for advanced behavioral demand models

Michel Bierlaire

Claudia Bongiovanni

Stefano Bortolomiol

École Polytechnique Fédérale de Lausanne (EPFL) **September 2021**

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

École Polytechnique Fédérale de Lausanne (EPFL)

Optimization algorithms for advanced behavioral demand models

Michel Bierlaire, Claudia Bongiovanni, Stefano Bortolomiol
Transport and Mobility Laboratory (TRANSP-OR)
École Polytechnique Fédérale de Lausanne (EPFL)
Station 18,1015 - Lausanne
phone:
fax:
{michel.bierlaire, claudia.bongiovanni, stefano.bortolomiol}@epfl.ch

September 2021

Abstract

Most real-world problems deal with decision-making at the individual level. Individual choices are typically represented through disaggregate demand models, which provide a realistic interpretation of human behavior by explicitly accounting for population heterogeneity. Disaggregate demand models have a substantial appeal for planning and design problems. However, due to their non-linear and non-convex nature, they require expensive estimation procedures based on simulation. Their disadvantage is therefore computational burden, which currently prohibits their use for realistically-sized decision-making problems.
Pacheco *et al.* (2021) proposed a simulation framework at the intersection of discrete choice and mathematical optimization, which allows to linearize disaggregate demand models and explicitly account for individual choices in mixed integer linear programs. The flexibility of this new framework comes at the price of computational complexity. Therefore, in order to speed-up the solution process, it is necessary to deploy mathematical decomposition algorithms that exploit the problems structure.
In this survey, we review choice-based optimization and mathematical decomposition, with the intention of combining the two in future research directions. Although the work is still at a conceptual stage, it is anticipated that this research direction will further demonstrate the power of combining techniques from the currently disconnected fields of optimization, discrete choice and simulation in solving crucial choice-based optimization problems.

Keywords

choice-based optimization, mathematical decomposition

1 Introduction

Most real-world problems in transportation and logistics are discrete decision-making problems that involve the choice of individuals, with different socio-economic characteristics and tastes, among a discrete and finite set of alternatives, that may need to be optimized. Such choice process is governed by complex random utility models, in which the probability that an individual picks a given alternative is in general a non-linear and non-convex function and has no closed analytical form (Manski (1977)). As a consequence of this complexity, random utility models have been typically: (i) estimated through extensive simulations on small-scale instances, and (ii) only explicitly considered for alternative optimization by relying on restrictive assumptions upon individual behavior.

Since the popularization of the logit model (McFadden (1974)), and thanks to the increased availability of customer-specific datasets, discrete choice models (DCM) based on the random utility principle have confirmed their power in predicting individual behavior. The main advantage of the logit model relies on its convex and closed-form probabilistic expression, which facilitates its direct embedding in choice-based optimization problems. However, the logit model is characterized by several assumptions that give rise to unrealistic estimates of individual behavior when alternatives are added to or deleted from the choice set, i.e. substitution patterns (Bierlaire (1998)). To relax such assumptions, more complex non-convex DCMs have been proposed in the literature, such as cross-nested logit models (e.g. Fosgerau *et al.* (2013)) and mixed logit models (e.g. McFadden and Train (2000)), to name a few. While advanced DCMs are better suited to predicting individual behavior in real-world settings, the complexity resulting from their non-convex mathematical form makes them unsuitable for large-scale decision-making problems.

In the literature, discrete decision-making problems involving discrete choice models, hereafter called "choice-based optimization" problems, have been tackled through optimization approaches that differ in their ability to produce high-quality solutions and to solve problems at scale. On the one hand, it is possible to assume simplistic discrete choice models (e.g. the logit) and optimize a unified choice-based assignment problem. This exact optimization approach can be applied to problems at scale but yields to approximated results as it relies on specific assumptions upon individual behavior (e.g. Alfandari *et al.* (2021), Akçakuş and Mišić (2021)). On the other hand, it is possible to rely on extensive simulations to linearize any non-convex advanced discrete choice model to optimize a unified choice-based assignment problem (Train (2009)). This other exact simulation-based optimization approach yields to optimal results but can only be applied to small-scale problem instances, as the optimization problem exponentially scales with the number of individuals, alternatives, as well as simulation replications (e.g. Pacheco (2020), Pacheco *et al.* (2021)). As such, new research directions are focusing on the design of appropriate mathematical decomposition frameworks to solve choice-based optimization

problems under no specific assumptions upon individual behavior and at scale (e.g. Pacheco *et al.* (2018), Bortolomiol *et al.* (2021)).
Matematical decomposition is an optimization field that aims at exploiting specific structural properties of decision-making problems to speed-up the solution process by parallelization (Conejo *et al.* (2006)). Discrete decision-making problems can be characterized by one or more of the following structural properties: (i) a set of complicating decision variables, and (ii) a set of complicating constraints (iii) an exponential number of decision variables. Each of these properties can be addressed through specific mathematical decomposition techniques. Namely, (i) Benders' decomposition (Benders (1962)), (ii) Lagrangian relaxation (Fisher (1981)), and (iii) column generation (Desaulniers *et al.* (2006)).
In this survey, we review choice-based optimization problems and mathematical decomposition techniques. In Section 2, we formally introduce the mathematical background for choice-based optimization, based on Pacheco *et al.* (2021), and provide some examples of transportation problems which are choice-based. In Section 3, we briefly review Benders' decomposition, Lagrangian relaxation, and column generation for a generic mathematical program which exhibits a structure that is common to multiple problems in transportation. Finally, Section 4 concludes this paper by examining choice-based problems that could benefit from mathematical decomposition and state possible challenges in future research.

2 Choice-based optimization

This Section formally introduces choice-based optimization in a simulation context, based on the work of Pacheco *et al.* (2021). For a comprehensive review on choice-based optimization for revenue management applications, the reader is referred to Strauss *et al.* (2018).
Consider a discrete and finite set of alternatives $i \in I$ which are offered to individuals $n \in N$. Note that set of available alternatives $\mathcal{I}$ also include the opt-out option. Each individual n is a choice maker whose goal is to maximize an utility function U_{in}, $\forall i \in I$. The explanatory variables of U_{in} include socio-economic characteristics and tastes of the individuals, as well as the attributes of the alternatives. The utility function U_{in} is of probabilistic nature and is decomposed into a systematic component V_{in}, which contains all of the observed variables, and a non-systematic component ϵ_{in}, which captures the randomness caused by unobserved variables and taste variations over time (Manski (1977)), as follows:

$$U_{in} = P_{in} = Pr[V_{in} + \epsilon_{in} = max_{j \in I} V_{in} + \epsilon_{in}]$$

As shown in Pacheco *et al.* (2021), the utility function U_{in} can be linearized via simulation. Specifically, we approximate the non-systematic part of U_{in} through $r = \{1, \ldots, R\}$ independent draws (or scenarios), for each alternative i and individual n. As such, the scenario-specific utility function can be re-written as:

$$U_{inr} = V_{in} + \zeta_{inr}$$

Where, ζ_{inr} is the drawn error term parameter for scenario r. Assuming individuals are utility maximizers, then the utility of the chosen alternative in scenario r is:

$$U^{max}_{nr} = \max_{j \in I} U_{jnr}$$

and the chosen alternative $x_{inr} = 1 \iff U_{inr} = U^{max}_{nr}$, for each individual n and scenario r. As such, assuming U_{inr} is given $\forall i \in \mathcal{I}$, $\forall n \in \mathcal{N}$, and $\forall r \in \mathcal{R}$, any choice-based optimization problem can be seen an optimization problem containing the following *knapsack problem*:

$$\max_{x_{inr}} \sum_{i \in \mathcal{I}} U_{inr} x_{inr}$$

s.t.

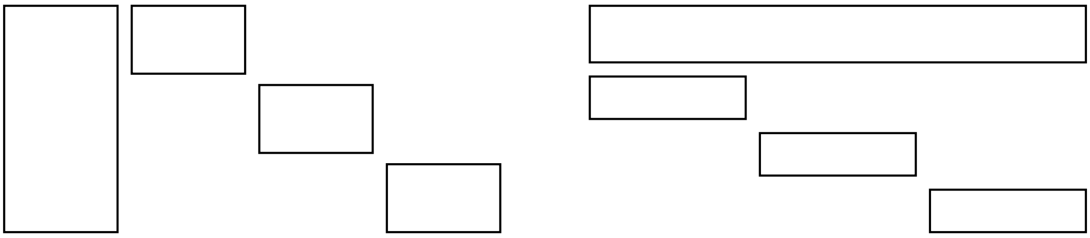
$$\sum_{i \in \mathcal{I}} x_{inr} = 1 \qquad \forall n \in \mathcal{N}, \forall r \in \mathcal{R} \tag{1}$$

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

Note that if all U_{inr} are different across i, then the optimal solution to (1) is binary.
Finally, there exists multiple transportation problems that are choice-based. Other than natural applications in revenue management (Strauss *et al.* (2018)), other relevant applications include: (i) network design (e.g. Farahani *et al.* (2013)), (ii) network pricing (e.g. Gilbert *et al.* (2014)), (iii) facility location (e.g. Haase and Müller (2014)), (iv) assortment optimization (Alfandari *et al.* (2021), Akçakuş and Mišić (2021)), (v) inventory control (e.g. Atzeni *et al.* (2012)), and (vi) train timetable design (Robenek *et al.* (2018)).

3 Mathematical decomposition

This Section introduces mathematical decomposition techniques that can be employed for choice-based combinatorial optimization problems characterized by the following structural properties: (i) a set of complicating decision variables, (ii) a set of complicating constraints, and (iii) an exponential number of decision variables. For each problem structure, we provide a brief overview on the specific mathematical decomposition technique for a general mathematical



(a) Decomposition: complicating variables (b) Decomposition: complicating constraints

program, based on the material in Gendron (2016). The reader is referred to Conejo *et al.* (2006) for a comprehensive review on mathematical decomposition.

3.1 A general mathematical program

Consider a general mathematical program with a linear objective function and linear constraints, in which part of the decision variables y are integer while others x are continuous, as follows:

$$Z(M) = \min f^T y + c^T x$$

s.t.

$$Ax = b$$

$$Bx + Dy \geq e \tag{2}$$

$$Gy \geq h$$

$$x, y \geq 0$$

$$y \in \mathbb{Z}^+$$

Such mathematical program can be characterized by block-diagonal constraints linking the integer decision variables y and the continuous decision variables x. Depending on the specific structure of such constraints, the mathematical program is either characterized by complicating variables (as depicted in Figure 1(a)) or complicating constraints (as depicted in Figure 1(b)). Examples of problems exhibiting the mathematical structure of (2) include network design (Farahani *et al.* (2013)), facility location (e.g. Haase and Müller (2014)), and assortment optimization (Alfandari *et al.* (2021), Akçakuş and Mišić (2021)).

3.2 Complicating variables: A review on Benders' decomposition

Combinatorial optimization problems that are characterized by complicating integer decision variables are typically tackled by a Benders' decomposition approach (Benders (1962)). The logic of this approach lies in the determination of the complicating integer variables of the mathematical problem, which can be temporarily fixed to give rise to much simpler linear subproblems to be solved. With this premise, a mathematical program is therefore split into two problems: (i) a problem containing all integer decision variables and constraints (i.e. the master problem), and (ii) a problem containing all continuous decision variables and constraints (i.e. the sub-problem). The master problem is solved, and an integer-feasible solution is found. This integer solution is successively used to solve the linear subproblem and produce: (i) feasibility cuts, if the subproblem is infeasible with the integer solution from the master problem; or (ii) optimality cuts, if the subproblem provides an optimal solution given the fixed integer solution from the master problem.. Note that if the master problem is found to be infeasible at any point, the overall problem is infeasible. Similarly, if the master problem is found to be unbounded at any point, the overall problem is unbounded. Examples of problems in transportation that are typically solved by Benders' decomposition include network design problems (e.g. Costa (2005)), facility location problems (e.g. Fischetti *et al.* (2017)), and resource management (Cai *et al.* (2001)), to name a few. For a review on Benders' decomposition, the reader is referred to Rahmaniani *et al.* (2017).

Let us consider the mathematical program (2), introduced in Section 3.1, featuring the complicating linking constraints as shown in Figure 1(a). The rationale of Benders' decomposition is based on the recognition of the complicating variables y, which can be fixed to feasible values $\bar{y} = \{y \in \mathbb{Z}^+ \mid Gy \geq h\}$, to successively solve the following simple linear program, called the Benders' subproblem:

$$
\begin{aligned}
& Z_x(\bar{y}) = \min c^T x \\
& \text{s.t.} \\
& Ax = b \\
& Bx \geq e - D\bar{y} \\
& x \geq 0
\end{aligned}
\tag{3}
$$

Note that the Benders' subproblem (3) can be re-written in its dual form as:

$Z_x(\bar{y}) = \max \pi b + \alpha(e - D\bar{y})$

s.t.

$$\pi A x + \alpha B \leq c \tag{4}$$

$\alpha \geq 0$

On the one hand, if the Benders' subproblem (3) is infeasible then its dual (4) is unbounded and there exists and extreme ray (π_j, α_j) of the dual polyhedron $D = \{(\pi, \alpha) \mid \pi A + \alpha B \leq c, \alpha \geq 0\}$ such that $\pi_j b + \alpha_j(e - D\bar{y}) > 0$. On the other hand, if the Benders' subproblem (3) is feasible then its dual (4) is also feasible and there exists an extreme point of the dual polyhedron D, (π_k, α_k) such that $\pi_k b + \alpha_k(e - D\bar{y}) = Z_x(\bar{y})$. With such premises, the original problem (2) can be re-written as follows:

$Z(M) = \min_{y} f^T y + Z_x(y)$

$Z_x(y) = \{\max_{k}(\pi_k b + \alpha_k(e - Dy)) \mid \pi_j b + \alpha_j(e - Dy) \leq 0, j \in J\}$

and by replacing $Z_x(y)$ by a variable z, we obtain:

$Z(M) = \min_{y,z} f^T y + z$

s.t.

$$\pi_k b + \alpha_k(e - Dy) \leq z \qquad \forall k \in K \tag{5}$$

$\pi_j b + \alpha_j(e - Dy) \leq 0 \qquad \forall j \in J$

Finally, we can solve the master problem (5) through the following cutting-plane procedure:

1. Start by solving a relaxed master problem with no constraints associated with K and J
2. At every iteration, solve the Benders' subproblem to derive:
   - an extreme point (π_k, α_k), or
   - an extreme ray (π_j, α_j)
3. If you find an extreme point (π_k, α_k), add the optimality cut: $\pi_k b + \alpha_k(e - Dy) \leq z$
4. If you find an extreme ray (π_j, α_j), add the feasibility cut: $\pi_j b + \alpha_j(e - Dy) \leq 0$
5. At every iteration, let $(\bar{y}, \bar{z})$ be the optimal solution to the relaxed master problem, and $\bar{x}$ be the optimal solution to the Benders' subproblem (if feasible), then: $f\bar{y} + \bar{z} \leq Z(M) \leq f\bar{y} + c\bar{x}$
6. The algorithm stops when $\bar{z} = c\bar{x}$

3.3 Complicating constraints: A review on Lagrangian relaxation

Combinatorial optimization problems featuring complicating constraints can be addressed by a Lagrangian relaxation approach (Fisher (2004)). The logic of this approach lies in the determination of the complicating constraints of the mathematical problem, which can be multiplied with Lagrangian dual variables and transferred into the objective function. This process produces a relaxed problem which is easier to solve and which provides optimistic bounds to the optimal objective value to the original problem. Examples of problems in transportation that are solved by Lagrangian relaxation include generalized assignment problems (e.g. Jörnsten and Näsberg (1986)), fleet sizing for network flow problems (e.g. Desrosiers *et al.* (1988)), location-routing (e.g.Shan *et al.* (2020)), and inventory routing (Chow and Nurumbetova (2015)). For a review on Lagrangian relaxation, the reader is referred to Fisher (2004).
Let us consider the mathematical program (2), introduced in Section 3.1, featuring the complicating linking constraints as shown in Figure 1(b). If we relax the complicating linking constraints $Bx + Dy \geq e$ using lagrangian multipliers α, we result in the following Lagrangian sub-problem:

$$Z(L(\alpha)) = \min_{x,y}(c - \alpha B)x + (f - \alpha D)y + \alpha e$$

s.t.

$$Ax = b$$

$$Gy \geq h$$

$x, y \geq 0$

$y \in \mathbb{Z}^+$

Note that, since $L(\alpha)$ is obtained by relaxing the complicating linking constraints between x and y, $Z(L(\alpha))$ is a lower bound on $Z(M)^*$ (i.e. $Z(L(\alpha)) \leq Z(M)^*$). Even if $Z(L(\alpha))$ is only a lower bound on the original problem, it allows us to define optimality gaps (i.e. to depict how far a given solution is from optimality). This is a very useful information in practice, as it allows us to assess the degree of sub-obtimality of a given solution and terminate our search early. In order to obtain the best bound on $Z(M)^*$ using $Z(L(\alpha))$, we solve the following Lagrangian dual problem:

$$Z(LD(M)) = \max_{\alpha \geq 0} Z(L(\alpha))$$

The most popular, since very easy to implement, choice to obtain optimal or near-optimal multipliers are subgradient algorithms (Boyd *et al.* (2003)). However, note that another compuation method based on the replacement of Y by $conv(Y)$ also exists (see Wolsey and Nemhauser (1999)). Note that $Z(L(\alpha))$ is continuous and concave but non-differentiable in α. A subgradient of $Z(L(\alpha))$ at $\bar{\alpha}$ is given by $(e - B\bar{x} - D\bar{y})$, where $(\bar{x}, \bar{y})$ solves the Lagrangian subproblem for $\alpha = \bar{\alpha}$. For a review on the subgradient method for lagrangian relaxation, the reader is referred to Section 3.5.3 in Pacheco (2020).

3.4 Exponential number of decision variables: A review on column generation

Combinatorial optimization problems featuring an exponential number of decision variables can be addressed through a column generation approach (Lübbecke (2010)). Such approach iteratively adds the decision variables from the mathematical model and employs duality theory to determine whether there are no more decision variables that can be added which would result in an improved solution (i.e. they have no negative reduced cost). In this case, an optimal solution has been found with only a subset of the decision variables and the algorithm can be terminated at an early stage (Lübbecke (2010)). This process is referred to as the pricing sub-problem and as such branch-and-bound algorithms that are enhanced by column generation are referred to as branch-and-price algorithms (Vanderbeck (2000)). Examples of problems in transportation that are typically solved by column generation include vehicle routing problems (e.g. Ceselli *et al.* (2009), Feillet (2010)), crew scheduling problems (e.g. Desaulniers *et al.* (2002)), and activity-based scheduling (Boyer *et al.* (2014)). For a review on column generation, the reader is referred to Desaulniers *et al.* (2006).

Let us consider the mathematical program (2), introduced in Section 3.1, featuring an exponential number of decision variables. If we ignore the continuous decision variables x and complicating constraints $Bx + Dy \geq e$, the remaining problem may correspond to a routing sub-problem, as for example a multi-commodity network flow problem (MCNFP) (e.g. Trivella *et al.* (2021)). In the MCNFP, a set of commodities $k \in \mathcal{K}$ are to be routed through a directed network $G = (N, A)$, with node set N and arc set A, with arc capacities u_{ij} and commodity-specific costs c^k_{ij}. The goal of the problem is to satisfy the commodity demand d_k between origins $O(k)$ and destinations $D(k)$ by using only one path per commodity at minimum cost, while respecting capacity constraints. As such, the MCNFP can be formulated as follows:

$$Z = \min \sum_{(i,j)\in\mathcal{A}} \sum_{k\in\mathcal{K}} c^k_{ij} y^k_{ij}$$

s.t.

$$\sum_{j\in N_i^+} y^k_{ji} - \sum_{j\in N_i^-} y^k_{ij} = \begin{cases} 1 & \text{if } i = O(k), \\ -1 & \text{if } i = D(k), \\ 0 & \text{otherwise.} \end{cases} \qquad \forall i \in \mathcal{N}, \forall k \in \mathcal{K}$$

$$\sum_{k\in K} y^k_{ij} \leq u_{ij} \qquad \forall (i,j) \in \mathcal{A}$$

$$y^k_{ij} \in \{0,1\} \qquad \forall (i,j) \in \mathcal{A}, \forall k \in \mathcal{K}$$

It is a well-known result in network flow theory that an extreme point $x_p = (x_{pij})$ of the polytope defined by the convex hull of $\mathcal{Y}$ corresponds to a circuit-free path $p \in \mathcal{P}^k$ between $O(k)$ and $D(k)$ for each k (Desaulniers *et al.* (2006)). This enables us to express the MCNFP as a convex combination of path flows and derive the following master problem:

$$Z = \min \sum_{k\in\mathcal{K}} \sum_{p\in\mathcal{P}^k} \Big(\sum_{(i,j)\in\mathcal{A}} c^k_{ij} \delta^{kp}_{ij} \Big) \lambda^{kp}$$

$$\sum_{p\in\mathcal{P}^k} \lambda^{kp} = 1 \qquad \forall k \in \mathcal{K} \qquad \rightarrow (\theta_k) \tag{6}$$

$$\sum_{k\in\mathcal{K}} \sum_{p\in\mathcal{P}^k} d^k \delta^{kp}_{ij} \lambda^{kp} \leq u_{ij} \qquad \forall (i,j) \in \mathcal{A} \qquad \rightarrow (\alpha_{ij})$$

$$\lambda^{kp} \geq 0 \qquad \forall k \in \mathcal{K}, \forall p \in \mathcal{P}^k$$

$$\delta^{kp}_{ij} \in \{0,1\} \qquad \forall (i,j) \in \mathcal{A}, \forall k \in \mathcal{K}, \forall p \in \mathcal{P}^k$$

where $\delta_{ij}^{kp} = 1 \iff (i, j) \in \mathcal{P}^k$, and 0 otherwise.
One starts with solving the linear programming (LP) relaxation of the path-based MCNFP. However, note that $\mathcal{P}^k$ generally has a high cardinality and ,consequently, the LP relaxation of the path-based model has an exponential number of decision variables. Column generation proposes a solution to this problem by generating a subset of sufficiently meaningful variables, forming the so-called restricted master problem (RMP) (Desaulniers *et al.* (2006)). More variables are added iteratively if needed. As in the simplex algorithm (Dantzig and Thapa (2006)), we need to identify promising variables to enter the basis. The identification is done by solving a pricing sub-problem through the following procedure: (i) optimize the RMP to determine the current optimal objective function value $\bar{z}$ and dual multipliers θ_k and α_{ij}, and (ii) find the variable λ_{kp} with minimum negative reduce cost:

$$\bar{c}_{kp} = \sum_{(i,j)\in\mathcal{A}} (c_{ij}^k + \alpha_{ij} d^k)\delta_{ij}^{kp} - \theta^k < 0$$

If there is no variable with negative reduce cost $\bar{c}_{kp}$, the column generation method has converged to the optimal LP relaxation. Finally, there exists multiple branching rules on λ_{kp} to efficiently solve the pricing sub-problem at every node of the branch-and-bound tree, which is at the heart of any column generation approach (e.g. Barnhart *et al.* (1994), Barnhart *et al.* (2000)). New approaches employing machine learning for branching have also recently emerged (e.g. Lodi and Zarpellon (2017), Balcan *et al.* (2018), Morabit *et al.* (2021)).

4 Conclusion

This survey reviews choice-based optimization and mathematical decomposition for a new research direction combining the two fields. Mathematical decomposition is an appealing direction for choice-based optimization, as most of the problems in this field exhibits complicating decision variables, complicating constraints, and an exponential number of decision variables. For example, Bortolomiol *et al.* (2021) is currently tackling an assortment and pricing problem, in which individual utilities (i.e. lower-level decisions) are explicitly taken into account for the determination of supplier decisions (i.e. upper-level decisions). This specific choice-based problem is tackled through a Benders' decomposition approach as it is characterized by complicating decision variables, which link lower-level and upper level decisions. As for the work in Pacheco (2020), the problem exponentially scales with the number of individuals, the number of supply alternatives, as well as the number of draws used to approximate individual utilities. Assuming an extension in which lower-level decisions are subject to upper-level capacity constraints, the problem could consider a Lagrangian approach to relax the complicating capacity limits. Other problems of interest that involve the interplay of choice-based optimization and mathematical decomposition include activity-based scheduling (e.g. Pougala *et al.* (2019)), personnel scheduling/re-scheduling problems, as well as simultaneous routing-scheduling problems (for example in the context of hyperloop operations). For all of such problems, a column generation approach has already shown promising results in related problems disregarding individual choices (e.g. Cordeau *et al.* (2001), Chow and Nurumbetova (2015), Kamran *et al.* (2020)). Finally, although this new research direction is still at a conceptual stage, the main challenge in our unified approach will lie in the identification of specific problem characteristics arising from the discrete choice part that we can exploit to enhance the mathematical decomposition part. Such problem characteristics may be of mathematical nature (e.g. interpretation of duality, variance reduction) as well as data-driven nature (e.g. socio-economic considerations).

5 References

Akçakuş, İ. and V. Mišić (2021) Exact logit-based product design, *Available at SSRN 3875986*.

Alfandari, L., A. Hassanzadeh and I. Ljubić (2021) An exact method for assortment optimization under the nested logit model, *European Journal of Operational Research*, **291** (3) 830–845.

Atzeni, I., L. G. Ordóñez, G. Scutari, D. P. Palomar and J. R. Fonollosa (2012) Demand-side management via distributed energy generation and storage optimization, *IEEE Transactions on Smart Grid*, **4** (2) 866–876.

Balcan, M.-F., T. Dick, T. Sandholm and E. Vitercik (2018) Learning to branch, paper presented at the *International conference on machine learning*, 344–353.

Barnhart, C., C. A. Hane, E. L. Johnson and G. Sigismondi (1994) A column generation and partitioning approach for multi-commodity flow problems, *Telecommunication Systems*, **3** (3) 239–258.

Barnhart, C., C. A. Hane and P. H. Vance (2000) Using branch-and-price-and-cut to solve origin-destination integer multicommodity flow problems, *Operations Research*, **48** (2) 318–326.

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

Bierlaire, M. (1998) Discrete choice models, in *Operations research and decision aid methodologies in traffic and transportation management*, 203–227, Springer.

Bortolomiol, S., V. Lurkin, M. Bierlaire and C. Bongiovanni (2021) Benders decomposition for choice-based optimization problems with discrete upper-level variables, paper presented at the *21st Swiss Transport Research Conference*, no. CONF.

Boyd, S., L. Xiao and A. Mutapcic (2003) Subgradient methods, *lecture notes of EE392o, Stanford University, Autumn Quarter*, **2004**, 2004–2005.

Boyer, V., B. Gendron and L.-M. Rousseau (2014) A branch-and-price algorithm for the multi-activity multi-task shift scheduling problem, *Journal of Scheduling*, **17** (2) 185–197.

Cai, X., D. C. McKinney, L. S. Lasdon and D. W. Watkins Jr (2001) Solving large nonconvex water resources management models using generalized benders decomposition, *Operations Research*, **49** (2) 235–245.

Ceselli, A., G. Righini and M. Salani (2009) A column generation algorithm for a rich vehicle-routing problem, *Transportation Science*, **43** (1) 56–69.

Chow, J. Y. and A. E. Nurumbetova (2015) A multi-day activity-based inventory routing model with space–time–needs constraints, *Transportmetrica A: Transport Science*, **11** (3) 243–269.

Conejo, A. J., E. Castillo, R. Minguez and R. Garcia-Bertrand (2006) *Decomposition techniques in mathematical programming: engineering and science applications*, Springer Science & Business Media.

Cordeau, J.-F., G. Stojković, F. Soumis and J. Desrosiers (2001) Benders decomposition for simultaneous aircraft routing and crew scheduling, *Transportation science*, **35** (4) 375–388.

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

Dantzig, G. B. and M. N. Thapa (2006) *Linear programming 1: introduction*, Springer Science & Business Media.

Desaulniers, G., J. Desrosiers and M. M. Solomon (2002) Accelerating strategies in column generation methods for vehicle routing and crew scheduling problems, in *Essays and surveys in metaheuristics*, 309–324, Springer.

Desaulniers, G., J. Desrosiers and M. M. Solomon (2006) *Column generation*, vol. 5, Springer Science & Business Media.

Desrosiers, J., M. Sauvé and F. Soumis (1988) Lagrangian relaxation methods for solving the minimum fleet size multiple traveling salesman problem with time windows, *Management Science*, **34** (8) 1005–1022.

Farahani, R. Z., E. Miandoabchi, W. Y. Szeto and H. Rashidi (2013) A review of urban transportation network design problems, *European Journal of Operational Research*, **229** (2) 281–302.

Feillet, D. (2010) A tutorial on column generation and branch-and-price for vehicle routing problems, *4or*, **8** (4) 407–424.

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

Fisher, M. L. (1981) The lagrangian relaxation method for solving integer programming problems, *Management science*, **27** (1) 1–18.

Fisher, M. L. (2004) The lagrangian relaxation method for solving integer programming problems, *Management science*, **50** (12_supplement) 1861–1871.

Fosgerau, M., D. McFadden and M. Bierlaire (2013) Choice probability generating functions, *Journal of Choice Modelling*, 1–18.

Gendron, B. (2016) Decomposition for network design, `https://transp-or.epfl.ch/courses/Gendron2016/index.php`.

Gilbert, F., P. Marcotte and G. Savard (2014) Logit network pricing, *Computers & operations research*, **41**, 291–298.

Haase, K. and S. Müller (2014) A comparison of linear reformulations for multinomial logit choice probabilities in facility location models, *European Journal of Operational Research*, **232** (3) 689–691.

Jörnsten, K. and M. Näsberg (1986) A new lagrangian relaxation approach to the generalized assignment problem, *European Journal of Operational Research*, **27** (3) 313–323.

Kamran, M. A., B. Karimi and N. Dellaert (2020) A column-generation-heuristic-based bendersâ decomposition for solving adaptive allocation scheduling of patients in operating rooms, *Computers & Industrial Engineering*, **148**, 106698.

Lodi, A. and G. Zarpellon (2017) On learning and branching: a survey, *Top*, **25** (2) 207–236.

Lübbecke, M. E. (2010) Column generation, *Wiley encyclopedia of operations research and management science. Wiley, New York*, 1–14.

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

McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior, paper presented at the *Frontiers in Econometrics*, 105–142, New York.

McFadden, D. and K. Train (2000) Mixed mnl models of discrete response, *Journal of Applied Econometrics*, **15**, 447–470.

Morabit, M., G. Desaulniers and A. Lodi (2021) Machine-learning–based column selection for column generation, *Transportation Science*, **55** (4) 815–831.

Pacheco, M. (2020) A general framework for the integration of complex choice models into mixed integer optimization, Ph.D. Thesis, École Polytechnique Fédérale de Lausanne.

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.

Pacheco, M., B. Gendron, V. Lurkin and S. S. A. M. Bierlaire (2018) A lagrangian relaxation technique for the demand-based benefit maximization problem, paper presented at the *Proceedings of the 18th Swiss Transport Research Conference (Ascona, Switzerland).*

Pougala, J., T. Hillel and M. Bierlaire (2019) Scheduling of daily activities: an optimization approach, paper presented at the *The 5th NYUAD Transportation Symposium*, no. POST_TALK.

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.

Robenek, T., S. S. Azadeh, Y. Maknoon, M. de Lapparent and M. Bierlaire (2018) Train timetable design under elastic passenger demand, *Transportation research Part b: methodological*, **111**, 19–38.

Shan, W., Z. Peng, J. Liu, B. Yao and B. Yu (2020) An exact algorithm for inland container transportation network design, *Transportation Research Part B: Methodological*, **135**, 41–82.

Strauss, A. K., R. Klein and C. Steinhardt (2018) A review of choice-based revenue management: Theory and methods, *European Journal of Operational Research*, **271** (2) 375–387.

Train, K. E. (2009) *Discrete choice methods with simulation*, Cambridge university press.

Trivella, A., F. Corman, D. F. Koza and D. Pisinger (2021) The multi-commodity network flow problem with soft transit time constraints: Application to liner shipping, *Transportation Research Part E: Logistics and Transportation Review*, **150**, 102342.

Vanderbeck, F. (2000) On dantzig-wolfe decomposition in integer programming and ways to perform branching in a branch-and-price algorithm, *Operations Research*, **48** (1) 111–128.

Wolsey, L. A. and G. L. Nemhauser (1999) *Integer and combinatorial optimization*, vol. 55, John Wiley & Sons.