An Outer-Inner Approximation Method for the Generic Choice-based Optimization Problem

Haoye Chen[a], Jan Kronqvist[b], and Zhenliang Ma[a,*]

[a] Department of Civil and Architectural Engineering, KTH Royal Institute of Technology, Stockholm, Sweden
haoye@kth.se, zhema@kth.se

[b] Department of Mathematics, KTH Royal Institute of Technology, Stockholm, Sweden
jankr@kth.se

[*] Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30), September 02-03, 2024, Crete, Greece

August 13, 2024

Keywords: Choice-based optimization, Outer-inner approximation, Multinomial logit model

1 Introduction

1.1 Background

Choice-based optimization combines demand modeling with optimization models for supply decisions. The key challenge of choice-based optimization is the nonlinearity of discrete choice models such as the multinomial logit model (MNL). Generally, three different categories of methods are proposed in the literature. The methods in Category I usually linearly reformulate the MNL model using techniques of variable substitution (Haase & Müller, 2014). Category II uses the unimodular constraint matrix to reformulate their choice-abased problem into a linear programming problem (Davis *et al.*, 2013). However, the methods in Categories I and II are constrained for specific problem structures, such as binary supply decisions and fixed attributes of options. Category III uses simulation-based sampling to approximate the MNL model constraints (Pacheco Paneque *et al.*, 2021) with significant computational challenges for a large number of sampling draws. Moreover, it is not applicable for continuous supply decisions because of its discretized sampling method. The paper proposes an outer-inner approximation method for the generic choice-based optimization problem with no specific problem structure requirement.

1.2 Problem definition

This subsection introduces the mathematical definition of the generic choice-based optimization problem. To model customer behavior, define W as the set of scenarios and R^w the choice set for scenario $w \in W$. For scenario w, μ^{wr} and θ^{wr} denote the deterministic utility and the chosen probability of option $r \in R^w$. The general formulation for the choice-based optimization P_0:

$$\min f(\boldsymbol{\theta}, \mathbf{x}), \tag{1}$$

$$\text{s.t.}\ \mathbf{s} = h(\boldsymbol{\theta}, \mathbf{x}), \tag{2}$$

$$\boldsymbol{\mu} = \mathbf{A}\mathbf{s}, \tag{3}$$

$$\theta^{wr} = \frac{\exp(\mu^{wr})}{\sum_{r' \in R_w} \exp(\mu^{wr'})}, \quad \forall r \in R^w,\ \forall w \in W, \tag{4}$$

$$\mathbf{x} \in \Omega, \tag{5}$$

where $\boldsymbol{\theta}$ is the customer behaviour, $\mathbf{x}$ are the operation decisions, and $\boldsymbol{\mu}$ is the utility of options. In objective (1), the general cost f is influenced by service operations $\mathbf{x}$ and customer behavior $\boldsymbol{\theta}$. Constraint (2) describes that the operation decisions and customer behavior determine service levels. Constraint (3) defines the utility of options as a linear function of operations and customer behavior where $\mathbf{A}$ is the coefficient matrix. Constraint (4) is MNL model for customer behavior. Constraint (5) defines the feasible space of service operations given the practical requirements.

2 Methodology

2.1 MINLP reformulation

The main computational challenge comes from the nonlinear nonconvex constraint (4). We focus on efficiently handling nonlinear constraints (4) in P_0, involving the ratio of choice probabilities of option pairs. Then, as introduced in Chen *et al.* (2023), we make the following assumptions:

1. θ^{wr} can only take values of $0 \cup [\epsilon, 1]$ where ϵ is a small threshold like 0.1%, that fulfills the accuracy requirements of the application. We call options with non-zero probabilities active options, otherwise, inactive options.
2. Consider the relationship (4) only for pairs of active options.
3. In comparison to active options, inactive options ought to exhibit inferior utilities, ensuring that they cannot satisfy the relationship (4) with a choice probability no less than ϵ.

We argue that these assumptions are reasonable. In practice, the original MNL model assigns exceptionally low probabilities to options, regardless of their inferiority as per constraint (4). Numerically, assumption 1 avoids the challenge of approximating $ln\theta$ when $\theta \to 0$. Assumption 2 ensures that all non-zero probabilities fulfill the MNL relationship. Assumption 3 ensures that options with 0 probability are unattractive options. Assumptions 2 and 3 avoid to compute the MNL relationship related to unattractive options. Thus, the assumptions exclude parts of the search space that are not interesting for the application but may create numerical challenges.

Binary variables $\boldsymbol{b} = \{b^{wr} | w \in W, r \in R^w\}$ are introduced that a value of 1 indicates an active route, 0 otherwise. We assume the absolute value of the utility μ has an upper bound $|U|_{max}$ and define continuous variables $\varphi^{wr} \in [\ln \epsilon - |U|_{max}, 0], w \in W, r \in R^w$. For scenario w, we select an arbitrary choice option r_0 as the base option. Then, the original problem P_0 can be represented by P within a bounded error (Chen *et al.*, 2023):

$$
\begin{aligned}
\mathrm{P}: \min_{\mathbf{x}} & f(\boldsymbol{\theta}, \mathbf{x}) \\
s.t.\ & (2), (3), (5), \\
& \theta^{wr} \in [0,1] && \forall w \in W,\ r \in R^w, && (6)\\
& \theta^{wr} \geq b^{wr}\epsilon && \forall w \in W,\ r \in R^w, && (7)\\
& \theta^{wr} \leq b^{wr} && \forall w \in W,\ r \in R^w, && (8)\\
& \varphi^{wr} - \varphi^{wr_0} = \mu^{wr} - \mu^{wr_0} && \forall w \in W, \forall r \in R^w/r_0, && (9)\\
& b^{wr} = 1 \rightarrow \varphi^{wr} = \ln \theta^{wr} && \forall w \in W, \forall r \in R^w, && (10)\\
& b^{wr} = 0 \rightarrow \varphi^{wr} < \ln \epsilon && \forall w \in W, \forall r \in R^w, && (11)\\
& b^{wr} \in \{0,1\}, \quad \ln \epsilon - |U|_{max} \leq \varphi^{wr} \leq 0 && \forall w \in W, \forall r \in R^w. && (12)
\end{aligned}
$$

Constraints (6), (7), and (8) restrict the ranges of θ^{wr}. They indicate that θ^{wr} is non-zero by $b^{wr} = 1$ and zero by $b^{wr} = 0$. Constraints (9) and (10) ensure assumption 2 holds when probabilities of two options are greater than ϵ and make natural log function starts from $\ln \epsilon$. Constraints (11) ensure assumption 3 holds. Constraints (12) define ranges of variables.

The nonlinearity in the reformulation comes from the service level function in constraint (2) and a concave nonlinear equality constraint (10) (its negative is convex). For simplicity, we introduce the equivalent form of problem P:

$$\tilde{\mathrm{P}}: \min f(\boldsymbol{\theta}, \mathbf{x}), \tag{13}$$
$$\text{s.t. } g(\mathbf{x}, \mathbf{s}, \boldsymbol{\theta}) = 0 \quad \forall g \in G_{\mathrm{P}}, \tag{14}$$
$$\mathbf{x}, \mathbf{s}, \boldsymbol{\mu}, \boldsymbol{\theta}, \mathbf{b} \in \boldsymbol{\Lambda}_{\mathrm{P}}, \tag{15}$$

where G_{P} is the set containing all nonlinear constraint functions in P and $\boldsymbol{\Lambda} = \{\mathbf{Bx} + \mathbf{Cs} + \mathbf{D}\boldsymbol{\mu} + \mathbf{E}\boldsymbol{\theta} + \mathbf{Fb} \leq 0\}$ is the solution space by all linear constraints. $\mathbf{B}, \mathbf{C}, \mathbf{D}$, $\mathbf{E}$, and $\mathbf{F}$ are coefficients matrices. We assume f and g are convex and once continuously differentiable. Note that the problem is a nonconvex MINLP due to the nonlinear equality constraints (14). Our approach for solving $\tilde{\mathrm{P}}$ builds upon the outer approximation approach (Duran & Grossmann, 1986).

2.2 Outer-inner Approximation method

We introduce projection $\mathrm{P}(\mathbf{b}_k)$ of P by fixing $\mathbf{b}$ to one assignment $\mathbf{b}_k$ and determining the optimal $\mathbf{x}$ variables for this assignment. R_k^w denotes the active options of scenario w for assignment $\mathbf{b}_k$. A feasible projection can provide an upper bound to P. The projection is given by

$$\begin{aligned}
\mathrm{P}(\mathbf{b}_k) : \min_{\mathbf{x}} & f(\boldsymbol{\theta}, \mathbf{x}) \\
s.t. \quad & (2), (3), (5), \\
& \theta^{wr} \in [\epsilon, 1] && \forall w \in W,\ r \in R_k^w, && (16) \\
& \theta^{wr} = 0 && \forall w \in W,\ r \in R^w / R_k^w, && (17) \\
& \varphi^{wr} - \varphi^{wr_0} = \mu^{wr} - \mu^{wr_0} && \forall w \in W, \forall r \in R_k^w / r_0, && (18) \\
& \varphi^{wr} = \ln \theta^{wr} && \forall w \in W, \forall r \in R_k^w, && (19) \\
& \ln \epsilon - |U|_{max} \leq \varphi^{wr} < \ln \epsilon && \forall w \in W, \forall r \in R^w / R_k^w. && (20)
\end{aligned}$$

However, the solution of $\mathrm{P}(\mathbf{b}_k)$ is not always feasible. We need an approach to exclude infeasible assignment $\mathbf{b}_k$. We define the feasibility problem $\mathrm{F}(\mathbf{b}_k)$ for infeasible $\mathrm{P}(\mathbf{b}_k)$:

$$\mathrm{F}(\mathbf{b}_k) : \min_{\mathbf{x}} \sum_{g \in G_{P(\mathbf{b}_k)}} |g(\mathbf{x}, \mathbf{s}, \boldsymbol{\theta})| \tag{21}$$
$$s.t. \quad \mathbf{x}, \mathbf{s}, \boldsymbol{\mu}, \boldsymbol{\theta} \in \boldsymbol{\Lambda}_{\mathrm{P}(\mathbf{b}_k)}. \tag{22}$$

Fletcher & Leyffer (1994) proved that the outer linearization at the solution $\mathbf{z}_k = (\mathbf{x}_k, \mathbf{s}_k, \boldsymbol{\theta}_k)^T$ of $\mathrm{F}(\mathbf{b}_k)$ can exclude $\mathbf{b}_k$ in $\boldsymbol{\Lambda}_{\mathrm{P}}$ when $g(\mathbf{z}_k) > 0$. As the functions g are convex we can use a piecewise linear approximation to outer approximate the equality constraint in the opposite direction, *i.e.,* $g(\mathbf{z}) \geq 0$. Thus, we can use classical gradient cuts for outer approximate $g(\mathbf{z}) \leq 0$ and a piecewise linear approximation to outer approximate $g(\mathbf{z}) \geq 0$. We prove that the piece-wise linearization (PWL) at the solution of $\mathrm{F}(\mathbf{b}_k)$ can exclude $\mathbf{b}_k$ in $\boldsymbol{\Lambda}_{\mathrm{P}}$ when $g(\mathbf{z}_k) < 0$. Therefore, we can define the master program M_i (for iteration i) containing outer linearizations (gradient cuts) and inner PWL for $g(\mathbf{z}) = 0$. The master problem is given by

$$\mathrm{M}_i : \min_{\mathbf{x}} \quad \eta \tag{23}$$
$$s.t. \quad \eta < \mathrm{UBD} \tag{24}$$
$$\eta \geq f_j(\mathbf{x}_j, \boldsymbol{\theta}_j) + (\nabla f_j)^{\mathrm{T}} \left((\mathbf{x}, \boldsymbol{\theta})^{\mathrm{T}} - (\mathbf{x}_j, \boldsymbol{\theta}_j)^{\mathrm{T}} \right) \quad \forall j \in T^i, \tag{25}$$
$$0 \geq g_j(\mathbf{x}_j, \boldsymbol{\theta}_j) + \left[\nabla g^j\right]^{\mathrm{T}} (\mathbf{z} - \mathbf{z}_j) \quad \forall g \in G, \forall j \in T^i, \tag{26}$$
$$0 \leq \mathbf{z} - \mathrm{PWL}\left(g(\mathbf{z}_l), g(\mathbf{z}_j), g(\mathbf{z}_r)\right) \quad \forall g \in G, \forall j \in T^i, \tag{27}$$
$$\mathbf{x}, \mathbf{s}, \boldsymbol{\mu}, \boldsymbol{\theta}, \mathbf{b} \in \boldsymbol{\Lambda}_{\mathrm{P}}, \tag{28}$$

and the solution provides a lower bound to the optimal objective value of P. Set $T^i = \{j | j <= i\}$

denotes previous and current iterations. $\mathbf{z} = (\mathbf{x}, \mathbf{s}, \boldsymbol{\theta})^{\mathrm{T}}$ denotes nonlinear-related variables. Given $\mathbf{b}_0$, $i = 0$, UBD= inf as initializations, the proposed algorithm can be summarized as follows: 1) Solve $\mathrm{P}(\mathbf{b}_i)$ or the feasibility problem $\mathrm{F}(\mathbf{b}_i)$ if $\mathrm{P}(\mathbf{b}_i)$ is infeasible, and let the solution be $\mathbf{z}_i = (\mathbf{x}_i, \mathbf{s}_i, \boldsymbol{\theta}_i)^T$. 2) Apply outer and inner PWL linearizations at $\mathbf{z}_i$ where $\mathbf{z}_l$ and $\mathbf{z}_r$ are the left and right boundaries of $\mathbf{z}$. 3) If $\mathrm{P}(\mathbf{b}_i)$ is feasible and f_i <UBD, record the current best solution $\mathbf{z}^* = \mathbf{z}_i$ and update UBD as f_i. 4) Solve the current relaxation M_i which produces a new assignment of options $\mathbf{b}_{i+1}$. 5) Move to next iteration $i + 1$ until M_i is infeasible.

3 Results

We conduct a preliminary case study of the network expansion problem to validate our proposed methodology. On the supply side, this problem aims to determine the extent of capacity expansion required for each road link. On the demand side, the travel patterns (choices) will change accordingly with expansion decisions. The objective is to reduce overall congestion by optimally expanding the capacities on certain links. We used the classical SiouxFalls network and OD demand data in `https://github.com/bstabler/TransportationNetworks/tree/master/SiouxFalls`.

For detailed settings, we treat OD pair w with travel demand d_w as one scenario and its available routes R^w as travel choice options. Set E contains all edges in the network. Decision variables $\{x_e | \underline{x}_e \leq x_e \leq \overline{x}_e, \forall e \in E\}$ denotes the adjusted capacities for all edges where, for edge e, $\underline{x}_e$ is the current road capacity and $\overline{x}_e$ is the maximum adjustable capacity. We set $\overline{x}_e = 1.5\underline{x}_e, \forall e \in E$. Objective (1) is specified as $f_1(\boldsymbol{\theta}) + f_2(\mathbf{x})$ in this problem. $f_1(\boldsymbol{\theta}) = \sum_w \sum_{r \in R^w} l_r d_w \theta^{wr}$ represents the total travel time within the system as the cost of customer behavior. l_r denotes the travel time associated with route r. $f_2(\mathbf{x}) = 0.005 \sum_{e \in E}(x_e - \underline{x}_e)$ quantifies the weighted cost of the operation decision (i.e., capacity expansion). We use BPR function as the service level (i.e., travel time) function and $\mu^{wr} = -1.5 l_r, \forall w \in W, r \in R^w$ for specified constraints (2) and (3). We tested on OD pairs $\{(i, j) | i < j, i, j <= 10\}$ for validation.

Table 1 shows the results by extending the capacities of 8 edges while retaining the other 68 edges. The problem $\mathbf{P}$ is solved in 17.96s with an i9-12900H CPU. The total travel time $f_1(\boldsymbol{\theta})$ decreases from 228,828 s to 222,240 s after the expansion (savings of 6587.7). The weighted expansion cost $f_2(\mathbf{x})$ is 127.28, significantly lower than savings. In the conference presentation, we will also show the case study of service frequency/pricing for multimodal mobility systems.

Table 1 – *The results of edges' capacities in the network*

Edge	$x_e/\underline{x}_e$	Edge	$x_e/\underline{x}_e$	Edge	$x_e/\underline{x}_e$
(2, 6)	150.00%	(5, 9)	150.00%	(9, 10)	148.49%
(4, 5)	107.62%	(6, 8)	150.00%	(16, 10)	150.00%
(5, 6)	150.00%	(8, 16)	150.00%	Remaning 68 edges	100.00%

References

Chen, Haoye, Kronqvist, Jan, Burghout, Wilco, Jenelius, Erik, & Ma, Zhenliang. 2023. Mixed Integer Formulation with Linear Constraints for Integrated Service Operations and Traveler Choices in Multimodal Mobility Systems. *In: hEART 2023, Zürich, Switzerland.*

Davis, James Mario, Gallego, Guillermo, & Topaloglu, Huseyin. 2013. Assortment Planning Under the Multinomial Logit Model with Totally Unimodular Constraint Structures.

Duran, Marco A, & Grossmann, Ignacio E. 1986. An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Mathematical programming*, **36**, 307–339.

Fletcher, Roger, & Leyffer, Sven. 1994. Solving mixed integer nonlinear programs by outer approximation. *Mathematical Programming*.

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

Pacheco Paneque, Meritxell, Bierlaire, Michel, Gendron, Bernard, & Sharif Azadeh, Shadi. 2021. Integrating advanced discrete choice models in mixed integer linear optimization. *Transportation Research Part B: Methodological*, **146**, 26–49.