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

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

[a] *Department of Civil and Architectural Engineering, KTH Royal Institute of Technology, Stockholm, 10044, Sweden*
[b] *Department of Mathematics, KTH Royal Institute of Technology, Stockholm, 10044, Sweden*

Abstract

Choice-based optimization problem integrates demand modeling into optimal supply decisions, which is generic for decision-making applications. Solving the problem is challenging given the nonlinear discrete choice model constraints. Existing solution methods are limited to specific problem structures, such as binary or discrete supply decisions and fixed option attributes. This paper proposes an outer-inner approximation method for the generic choice-based optimization problem without specific problem structural requirements. We validated the method using a network expansion problem on the SiouxFalls network, aiming to reduce the overall system congestion by optimally expanding road capacities considering the road expansion cost. The results show that the expansion cost is significantly lower than the total travel time savings. More experiments are expected to benchmark with existing models using more case studies, e.g., service frequency and pricing in multimodal transportation systems.

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

---

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 and 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

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. (2025), 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 option, 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., 2025):

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

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:

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

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 (Kronqvist et al., 2019) due to the nonlinear equality constraints (14). Our approach for solving $\tilde{\mathrm{P}}$ builds upon the outer approximation approach (Duran and 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 and 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

$$\mathrm{M}_i : \min_{\mathbf{x}} \quad \eta \tag{23}$$

$$s.t. \quad \eta < \mathrm{UBD} \tag{24}$$

$$\eta \geq f\left(\mathbf{x}_j, \boldsymbol{\theta}_j\right) + \left(\nabla f_j\right)^{\mathrm{T}} \left(\left(\mathbf{x}, \boldsymbol{\theta}\right)^{\mathrm{T}} - \left(\mathbf{x}_j, \boldsymbol{\theta}_j\right)^{\mathrm{T}}\right) \qquad \forall j \in T^i, \tag{25}$$

$$0 \geq g\left(\mathbf{z}_j\right) + \left[\nabla g_j\right]^{\mathrm{T}} \left(\mathbf{z} - \mathbf{z}_j\right) \qquad \forall g \in G, \forall j \in T^i, \tag{26}$$

$$0 \leq \mathbf{z} - \mathrm{PWL}\left(g(\mathbf{z}_l), g(\mathbf{z}_j), g(\mathbf{z}_r)\right) \qquad \forall g \in G, \forall j \in T^i, \tag{27}$$

$$\mathbf{x}, \mathbf{s}, \boldsymbol{\mu}, \boldsymbol{\theta}, \mathbf{b} \in \boldsymbol{\Lambda}_{\mathrm{P}}, \tag{28}$$

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 P($\mathbf{b}_i$) or the feasibility problem F($\mathbf{b}_i$) if 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 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. Algorithm 1 illustrates the above steps.

Algorithm 1: Outer-Inner Approximation

Given $\mathbf{b}_0$, $i = 0$, UBD= inf
$i \leftarrow 0$
while M_i *is feasible* **do**
    Solve $\mathrm{P}(\mathbf{b}_i)$
    if $\mathrm{P}(\mathbf{b}_i)$ *is feasible* **then**
        Record the solution as $(\mathbf{x}_i, \mathbf{s}_i, \boldsymbol{\theta}_i)^T$
    else
        Solve $\mathrm{F}(\mathbf{b}_i)$
        Record the solution as $(\mathbf{x}_i, \mathbf{s}_i, \boldsymbol{\theta}_i)^T$
    end
    $\mathbf{z}_i \leftarrow (\mathbf{x}_i, \mathbf{s}_i, \boldsymbol{\theta}_i)^T$
    Add outer linearization to M_i at $\mathbf{x}_i, \theta_i$ for f
    Add outer linearizations to M_i at $\mathbf{z}_i, \forall g \in G$
    Add inner PWL linearizations to M_i at $\mathbf{z}_i, \forall g \in G$
    $T^i \leftarrow T^{i-1} \cup i$
    if $\mathrm{P}(\mathbf{b}_i)$ *is feasible* **and** $f^i <$ UBD **then**
        Update the current best solution $\mathbf{z}^* = \mathbf{z}_i$
        Update UBD=f^i
    end
    Solve M_i, producing a new integer assignment $\mathbf{b}_i$
    $i \leftarrow i + 1$
end

3. Convergence Properties

As the problem is non-convex, we cannot guarantee convergence by solving the projection problem. This can result in cycling (getting the same solutions and integer assignment $\mathbf{b}_i$). The projection problem is mainly intended to speed up convergence.

To guarantee convergence, we always update the linear model at the solution of the master problem if we get the same integer assignment (same value for the binary variables as in an earlier iteration). We prove the convergence of the technique for avoiding cycling.

For simplicity, we consider the problem to be written in the form

$$\tilde{\mathrm{P}}: \min f(\boldsymbol{\theta}, \mathbf{x}), \tag{29}$$
$$\text{s.t. } g_i(x_i) = \theta_i \quad \forall i \in \mathcal{I}, \tag{30}$$
$$\mathbf{x}, \mathbf{s}, \boldsymbol{\mu}, \boldsymbol{\theta}, \mathbf{b} \in \boldsymbol{\Lambda}_{\mathrm{P}}, \tag{31}$$

where Λ_{P} is defined by linear constraints and integer restrictions on some variables.

For proving converge we need the following assumptions

1. The objective function f is convex and continuously differentiable.
2. Set Λ_{P} is compact. Let $\mathcal{X} \subset \mathbb{R}^n$ be the box defined by the x variables upper (ub) and lower (lb) bounds.
3. Functions g_i are concave and continuously differentiable. We focus on concave functions, but all the results trivially extend to convex functions.
4. Functions g_i are Lipschitz continuous with constant $L < \infty$ such that

$$|g(x^1) - g(x^2)| \leq L|x^1 - x^2| \quad \forall x^1, x^2 \in \mathcal{X} \tag{32}$$

   where $\mathcal{X}$ is defined by the variables' upper and lower bounds. We assume this holds for all functions g_i

   - For the problems considered here, this assumption clearly holds as the functions are $ln(x_i)$, and we only consider $x_i \in [\epsilon, ub]$ were ϵ is a positive user-defined parameter. In fact, we can set $L = ln(\epsilon)$.

Remember we consider $g_i(x_i,) = \theta_i$ as two inequality constraints. The first constraint is nonconvex and given by $g_i(x_i,) \leq \theta_i$, where g_i is a concave function. The second constraint is convex and given by $g_i(x_i,) \geq \theta_i$. Note that the only difference between the functions g_i would be that the first constraint would be convex and the second nonconvex.

We will focus on the nonconvex constraints $g_i(x_i,) \leq \theta_i$ as it is well known that outer approximation type algorithms can guarantee convergence for this class of problems without the nonconvex constraints (Duran and Grossmann, 1986; Fletcher and Leyffer, 1994). To show that the algorithm converges, we will show that the master problem forms a relaxation of the original problem, show that the algorithm ensures a sufficient improvement of the outer approximation of the feasible, and finally prove that the algorithm converges towards a feasible and optimal solution.

The algorithm uses a "fallback strategy" that differs from traditional outer approximation but is needed to ensure sufficient progress in this nonconvex setting. If the algorithm returns the same integer assignment in interaction i as in any previous iteration, then we refine the approximation of the linear approximation of the constraints $g_i(x_i,) \leq \theta_i$ and $g_i(x_i,) \geq \theta_i$ at the solution returned by the master problem. If a convex constraint is violated, we generate a gradient cut at the solution given by the master problem just as in the extended cutting plane algorithm (Westerlund and Pettersson, 1995; Eronen et al., 2014). If a nonconvex constraint $g_i(x_i,) \leq \theta_i$ is violated, we add a breakpoint to the piecewise linear approximation of the function $g_i(x_i,)$ at the value of x_i that is optimal for the master problem.

Proposition 1. *The master problem* M_i *forms a relaxation of problem* $\tilde{\mathrm{P}}$.

Proof. The gradient cuts clearly form an outer approximation of the feasible set of the convex constraints. Using piecewise linear approximations of the functions g_i in the constraints $g_i(x_i,) \leq \theta_i$ also result in an outer approximation as the functions g_i are concave. □

As a consequence of Proposition 1, we know that if the master problem returns a solution that satisfies all constraints within the accepted tolerances, then the solution is considered to be optimal. From here on, we will focus the convergence analysis on the "fallback" strategy. Because after a certain (possibly huge) number of iterations, the updates will always follow the fallback strategy as there is only a finite number of different integer assignments in Λ_{P}. If the master problem returns a solution satisfying the constraints, before exhausting all integer assignments the search can be terminated as an optimal solution has been found.

Let $\epsilon_g > 0$ denote an acceptable constraint tolerance for the constraints $g_i(x_i) = \theta_i$, *i.e.*, a solution satisfying the constraints within $\pm\epsilon_g$ is considered acceptable. Next, we analyze the improvement of the master problem in the case that at least one of the nonconvex constraints is violated.

Lemma 1. *Assume the master problem returns a solution* $\mathbf{x}^k, \theta^k$ *that violates at least one of the nonconvex constraints* $g_i(x_i,) \leq \theta_i$ *by* $\Delta^k > 0$. *Furthermore, assume the integer assignment returned by the master problem has already been explored resulting in the fallback updating strategy. Then, the updated piecewise linear approximation will ensure that any future solution*

to the master problem satisfies
$\| \begin{bmatrix} \mathbf{x} - \mathbf{x}^k \\ \theta - \theta^k \end{bmatrix} \|_1 \geq \min\{\frac{\Delta^k}{L}, \Delta^k\}$.

Proof. Follows from the Lipschitz continuity and the update of the piecewise linear approximation. The slope of all parts of the piecewise linear approximation of g_i will be less than or equal to the Lipschitz constant L. Therefore, the x variables need to change by at least $\frac{\Delta^k}{L}$ to satisfy the constraints in the master problem if the θ variables are kept fixed. The θ variables on the other hand need to change by Δ^k for $\mathbf{x}^k$ to be feasible in the update master problem. □

Now, we are ready to analyze the convergence of the solutions returned by the master problem. Assume the algorithms creates the infinite sequence of solutions $\{\mathbf{x}^i, \mathbf{s}^i, \boldsymbol{\mu}^i, \boldsymbol{\theta}^i, \mathbf{b}^i\}_{i=1}^{\infty}$. Next, we will analyze this sequence and show that the algorithm converges towards a feasible and optimal solution.

Proposition 2. *The sequence* $\{\mathbf{x}^i, \mathbf{s}^i, \boldsymbol{\mu}^i, \boldsymbol{\theta}^i, \mathbf{b}^i\}_{i=1}^{\infty}$ *contains at least one accumulation point* $(\bar{\mathbf{x}}, \bar{\mathbf{s}}, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\theta}}, \bar{\mathbf{b}})$.

Proof. This property is clear as each $(\mathbf{x}^i, \mathbf{s}^i, \boldsymbol{\mu}^i, \boldsymbol{\theta}^i, \mathbf{b}^i) \in \Lambda_{\mathrm{P}}$, and Λ_{P} is a compact subset of $\mathbb{R}^N$. □

Lemma 2. *An accumulation point* $(\bar{\mathbf{x}}, \bar{\mathbf{s}}, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\theta}}, \bar{\mathbf{b}})$ *of the sequence* $\{\mathbf{x}^i, \mathbf{s}^i, \boldsymbol{\mu}^i, \boldsymbol{\theta}^i, \mathbf{b}^i\}_{i=1}^{\infty}$ *is feasible.*

Proof. From Lemma 3.3 in (Eronen et al., 2014) we know that the accumulation point must satisfy the convex constraints. From Lemma 1, it is clear that the accumulation point cannot violate the nonconvex constraints. □

The main result is summarized in the following theorem.

Theorem 1. *Given a constraint tolerance* $\epsilon_g > 0$*, the algorithm terminates after a finite number of iterations and with an optimal solution within the given tolerance.*

Proof. Note that each point of the sequence $\{\mathbf{x}^i, \mathbf{s}^i, \boldsymbol{\mu}^i, \boldsymbol{\theta}^i, \mathbf{b}^i\}_{i=1}^{\infty}$ is a minimizer of a relaxation of the original problem. Due to Lemma 2 and Proposition 1, an accumulation point will be feasible and globally optimal. Within a finite number of iterations we can obtain a solution $(\mathbf{x}^k, \mathbf{s}^k, \boldsymbol{\mu}^k, \boldsymbol{\theta}^k, \mathbf{b}^k)$ arbitrarily close to the accumulation point $(\bar{\mathbf{x}}, \bar{\mathbf{s}}, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\theta}}, \bar{\mathbf{b}})$. □

4. 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). For edge $e \in E$, the associated travel flow is:

$$q_e = \sum_w \sum_{r \in R^w} d_w \theta^{wr} \delta_e^{wr}, \tag{33}$$

where δ_e^{wr} is an indicator that 1 if edge e belongs to the route $r \in R^w$. We use the BPR function as the service level (i.e., travel time) function.

The travel time for edge e:

$$t_e = t_0(1 + \alpha(q_e/x_e)^\beta), \tag{34}$$

where parameters $\alpha = 0.15$ and $\beta = 4$. It corresponds to constraint (2). Note that both q_e and x_e are variables. We define auxiliary variables $y_e = q_e/x_e, \forall e \in E$ and directly linearize y_e in the experiment.

The utility of option/route $r \in R^w$:

$$\mu^{wr} = -1.5 l_r, \forall w \in W, r \in R^w, \tag{35}$$

where $l_r = \sum_{e \in E} t_e \delta_e^{wr}$. It corresponds to constraint (3). We tested on OD pairs $\{(i,j) | i < j, i, j <= 10\}$ for validation.

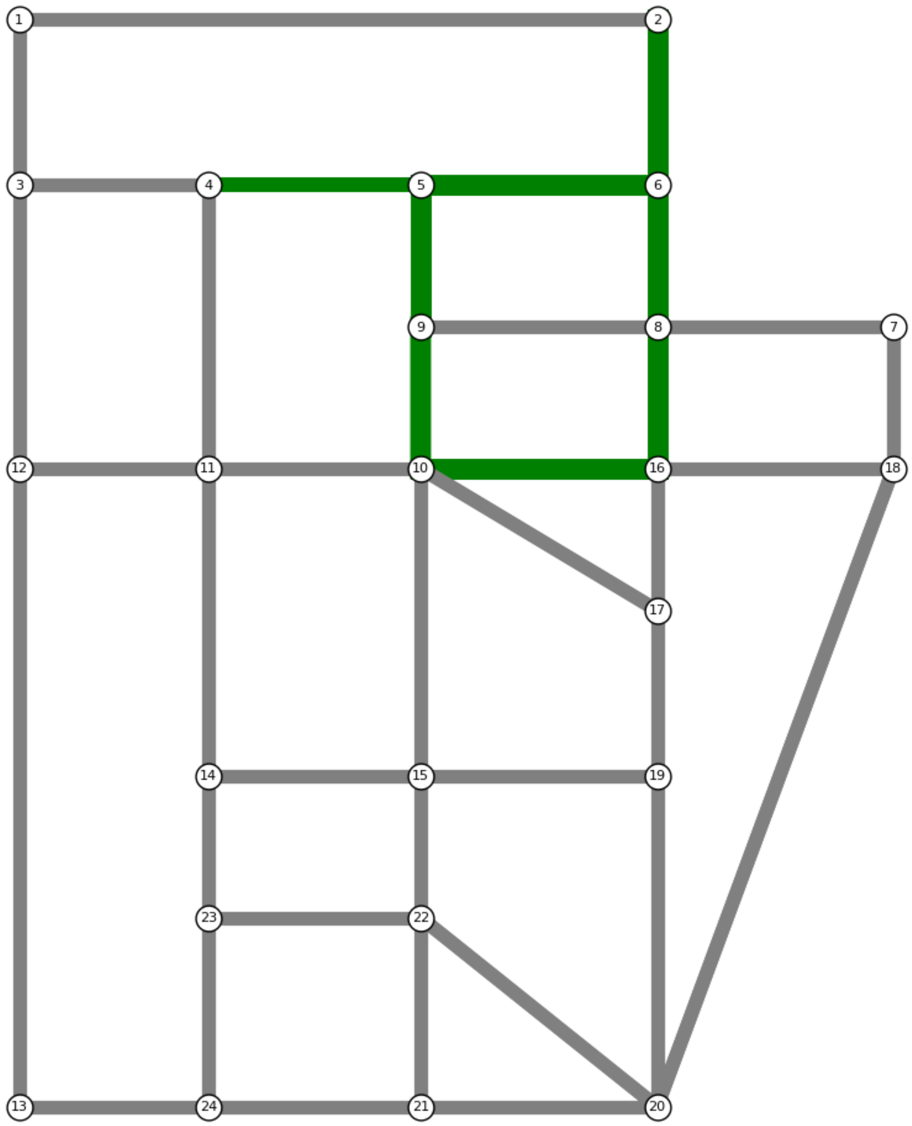

Figure 1: The expanded edges in the SiouxFalls network

In the experiment, the method takes 4 iterations to converge to the solution in which the master relaxation is no longer feasible (i.e., M_3 is infeasible). The results for the previous master relaxation are: $\mathrm{M}_0 = 220,630, \mathrm{M}_1 = 221,237,$ and $\mathrm{M}_2 = 221,805$,. The projection problem $\mathrm{P}(\mathbf{b}_i)$ is infeasible for $i \in \{0,1,2\}$. $\mathrm{P}(\mathbf{b}_3) = 222,368$ is the obtained solution.

Table 1 and Figure 1 show 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.

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%

5. Conclusion

We propose an outer-inner approximation to solve the generic choice-based optimization problem. The proposed method reformulates the problem into a MINLP formulation by only considering active options. Then, we propose an iterative algorithm updating the problem's upper bound through a projection based on an integer assignment. For the lower bound, we iteratively refine a master relaxation problem by applying outer approximation and inner piece-wise linearization for nonlinear constraints at guidance points obtained from projection or feasibility problems. We prove the convergence properties for the cycling case of identical integer assignment. The case study using network expansion validates the model performance in convergence and solution quality. We are testing the methods for more use cases, including location planning and pricing problems.

References

Chen, H., Kronqvist, J., Ma, Z., 2025. A choice-based optimization approach for service operations in multimodal mobility systems. Transportation Research Part C: Emerging Technologies 171, 104954. URL: `https://www.sciencedirect.com/science/article/pii/S0968090X24004753`, doi:`https://doi.org/10.1016/j.trc.2024.104954`.

Davis, J.M., Gallego, G., Topaloglu, H., 2013. Assortment planning under the multinomial logit model with totally unimodular constraint structures.

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

Eronen, V.P., Mäkelä, M.M., Westerlund, T., 2014. On the generalization of ecp and oa methods to nonsmooth convex minlp problems. Optimization 63, 1057–1073.

Fletcher, R., Leyffer, S., 1994. Solving mixed integer nonlinear programs by outer approximation. Mathematical Programming .

Haase, K., Müller, S., 2014. A comparison of linear reformulations for multinomial logit choice probabilities in facility location models. European Journal of Operational Research 232, 689–691. URL: https://www.sciencedirect.com/science/article/pii/S0377221713006747, doi:https://doi.org/10.1016/j.ejor.2013.08.009.

Kronqvist, J., Bernal, D.E., Lundell, A., Grossmann, I.E., 2019. A review and comparison of solvers for convex minlp. Optimization and Engineering 20, 397–455.

Pacheco Paneque, M., Bierlaire, M., Gendron, B., Sharif Azadeh, S., 2021. Integrating advanced discrete choice models in mixed integer linear optimization. Transportation Research Part B: Methodological 146, 26–49. URL: https://linkinghub.elsevier.com/retrieve/pii/S0191261521000266, doi:10.1016/j.trb.2021.02.003.

Westerlund, T., Pettersson, F., 1995. An extended cutting plane method for solving convex minlp problems. Computers & chemical engineering 19, 131–136.