Wiley
Journal of Advanced Transportation
Volume 2024, Article ID 8131887, 15 pages
https://doi.org/10.1155/2024/8131887

WILEY

Research Article

Optimizing Transit Network Departure Frequency considering Congestion Effects

Wei Tan,[1,2] **Xiaodong Peng**,[2] **Jun Huang**,[1] **Yuwen Wang**,[1] **Jiandong Qiu**,[3] **and Xiaobo Liu**[1]

[1]*School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, Sichuan 610031, China*
[2]*Southwest Municipal Engineering Design & Research Institute of China, Chengdu, Sichuan 610081, China*
[3]*Shenzhen Urban Transport Planning Center Co., Shenzhen 518000, China*

Correspondence should be addressed to Jun Huang; jun.huang@my.swjtu.edu.cn

Received 15 November 2023; Revised 16 April 2024; Accepted 4 May 2024; Published 28 May 2024

Academic Editor: Indrajit Ghosh

This paper introduces a bilevel programming model for optimizing transit network departure frequency. In the upper-level model, user satisfaction is reflected by considering congestion effects in the cost function. The lower-level assignment model simulates passenger travel behavior more realistically by incorporating congestion effects. This problem is solved by a heuristic gradient descent algorithm, where an approximation of the gradient is obtained at each iteration by using sensitivity analysis for transit equilibrium problems. The effectiveness of the proposed model and algorithm is demonstrated through two test examples, one of which involves a real-world scenario comprising over 130 transit lines. Numerical results conclusively indicate that the incorporation of congestion effects in the proposed model leads to improved transit system performance and enhanced user satisfaction.

1. Introduction

As a crucial part of urban transportation planning, transit system planning directly impacts the operational efficiency of urban transportation. The literature [1] identifies five stages in transit system planning: route network design, frequency setting, timetable design, fleet assignment, and crew assignment. Among these elements, transit frequency design plays a vital role in urban transit systems as different frequencies may lead to distinct vehicle schedules and driver assignments [2], which pose a significant obstacle for public transit agencies. This study will primarily focus on the transit frequency setting problem, which aims to determine the optimal line frequency for an existing transit network based on given passenger demand.

Some research is devoted to constructing single-layer analytical models for frequency optimization. Schéele [3] introduced a nonlinear programming model that considers the distribution of trips across different zones, aiming to minimize passenger travel times while implicitly incorporating user behavior. Furth and Wilson [4] proposed a frequency-setting problem that aims to maximize the social benefits of the transit system under constraints such as total subsidies, fleet size, and vehicle loading levels. Han and Wilson [5] developed an optimization model for heavily utilized lines, seeking to minimize passenger waiting times and bus congestion while operating under limitations in fleet size and bus capacity. The assignment submodel is expressed by an implicit constraint. Ceder [6] delineated several suitable data collection methods and proposed four approaches based on maximum load data and load profile data for adjusting bus route frequencies. Capali and Ceylan [7] utilized the intelligent water drops algorithm to solve the bus network design and frequency setting (BNDFS) problem. Ahern et al. [8] proposed a multiobjective optimization model for BNDFS in public transit systems. This model integrates service frequency and passenger assignment to reflect the available options for passengers in the network. Durán-Micco and Vansteenwegen [9] conducted a comprehensive review of the literature related to the BNDFS problem, focusing on papers from the past decade. They highlighted different problem definitions and assumptions.

Yoo et al. [10] introduced a novel approach to address the BNDFS problem. Even though this research has produced valuable models and solution algorithms, their single-layer nature often fails to adequately consider passenger choice behavior and the interactions between transit operators and passengers. Therefore, the development of more practical models and algorithms is imperative in this regard.

In a typical transit system, transit operators usually seek to optimize some particular criteria, such as the total passenger travel time, to determine the frequencies of transit lines. Accordingly, the passengers make their route choices in response to the decisions of the operators as followers with the goal of minimizing their travel costs. Evidently, the frequency setting problem exhibits the leader-follower model structure, and it is often treated as a bilevel optimization problem. Frequency optimization was first represented as a nonlinear bilevel problem by Constantin and Florian [11] where the upper- and lower-level objective functions both aim to reduce the total travel time. Gao et al. [12] proposed a multiobjective bilevel model that seeks to minimize total passenger travel time and operating costs. A notable characteristic of this work is the consideration of passenger travel behavior under congestion. Yu et al. [13] designed a bilevel optimization problem that aimed to minimize boarding and waiting times under fleet size constraints. It utilized an optimal strategies assignment model [14] to account for the route choice behavior of the users. Dell'Olio et al. [15] presented a constrained bilevel optimization model that allows for the allocation of buses of varying sizes to transit routes. The upper-level model considered the minimization of social and operating costs of the transit system, while the lower-level model represented the assignment model, imposing constraints on vehicle capacities. Martínez et al. [16] extended the work of Constantin and Florian [11] to determine the time intervals between subsequent buses on a set of transit lines and introduced mixed integer linear programming formulations. A novel metaheuristic approach was employed to solve the problem. Overall, while various frequency optimization models have been proposed in the literature, they share very similar objective functions and constraints. Typically, these functions aim to minimize user travel cost and/or operating cost under constraints such as fleet size and other infrastructure considerations. Since optimization models require measures of transit system performance from the user's perspective, they frequently incorporate a submodel of the route selection behavior of the users based on frequency, also known as the assignment submodel.

Although the frequency setting problem and its variants have been extensively studied in the literature, it is worth noting that there is a noticeable paucity of studies that take into account user satisfaction within these models [17]. Mo et al. [18] addressed the frequency optimization problem considering user satisfaction, aiming to schedule buses within a given waiting time threshold to serve more passengers, rather than minimizing the travel cost of passengers. To our knowledge, they were the first to emphasize user satisfaction in the frequency optimization problem. Existing research has demonstrated that passenger user satisfaction is typically influenced by riding comfort [19, 20], which is further impacted by vehicle congestion [21]. Transit vehicles also suffer significantly from the consequences of congestion effects in some large urban areas [22]. Therefore, it is necessary to consider congestion effects in the assignment submodel to reflect their influence on passenger route choice behavior. However, existing research often inadequately accounts for congestion effects when describing passenger route choice. Notable models in this regard include the optimal strategies assignment model [14], which neglects vehicle congestion, and the assignment model proposed by de Cea and Fernández [23], which only considers delay-waiting time due to capacity constraints. Wu et al. [22] and Xu et al. [21] extended the previous studies to model waiting and in-vehicle cost as functions of transit flows, thus effectively accounting for the impact of congestion effects including queuing and crowding effects on route choice behavior. In this study, we incorporate these congestion effects into the assignment submodel, aiming to create a more precise prediction of passenger travel behavior. Specifically, the main contributions of this paper are as follows.

First, instead of focusing solely on passenger travel time, we emphasize user satisfaction by incorporating cost functions that consider more realistic congestion effects in the context of transit frequency optimization. The objective of the model is to minimize a weighted sum of passenger travel cost (including congestion cost) and operational cost.

Second, we consider the influence of congestion effects (including queuing and crowding effects) on passenger travel behavior within the assignment submodel. This allows for a more realistic representation of changes in passenger travel patterns under different frequencies. Furthermore, congestion effects arise from the crowding effect, queuing effect, capacity effect, and bunching effect according to the literature [21]. In this paper, we mainly focus on queuing and crowding effects. While considering these effects correctly characterizes the route choice behavior, it also results in asymmetric cost functions, thereby posing computational challenges [21]. Furthermore, due to the nature of bilevel problems, the complexity of the assignment submodel is a critical component influencing the overall complexity of the frequency optimization. Thus, there is a need to seek efficient solution algorithms for the assignment submodel. In this work, we employ the gradient projection-adaptive inner looping (GP-AIL) algorithm proposed by Xu et al. [21] to solve it, and it demonstrates the capability to converge to solutions within a relatively short timeframe, even in large-scale networks.

Third, we introduce a heuristic gradient descent algorithm to effectively address the frequency optimization. Since frequency optimization is typically an NP-hard problem, traditional optimization techniques often exhibit lower computational efficiency when solving the model mentioned above, especially in the case of large-scale networks [13]. Furthermore, there is a lack of methods in the existing literature that are capable of finding globally optimal solutions [16]. Thus, in recent years, many studies have applied heuristic descent algorithms to frequency

optimization [11, 12]. These successful results have motivated us to employ this type of algorithm in this work. Numerical experiments indicate that the algorithm can converge to sufficiently close local solutions within a relatively short timeframe.

The structure of the paper is as follows. Section 2 introduces the network representation and assignment model based on hyperpaths, and a bilevel programming model for frequency optimization and its solution algorithm is presented in Section 3. Section 4 provides numerical results and analytical conclusions for various test cases. Finally, Section 5 offers remarks on the conclusions and outlines directions for future work.

2. Network Representation and Assignment Model

2.1. Transit Network Representation and Hyperpaths. Consider a transit network composed of a series of distinct transit lines, transit stops, and centroid nodes generating and attracting travel demand. This network is abstracted, as depicted in Figure 1, where O and D represent the origin and destination, respectively, and a typical transit stop n_1 is served by multiple transit lines. While Figure 1 can illustrate the layout of transit lines by describing their alignments, it may not effectively reflect the travel behavior of passengers. Scholars [14, 23] often expand the transit network to provide a more comprehensive representation of passenger travel behavior, including transfer, waiting, boarding, and alighting behavior.

We define $\mathcal{G} = (N, A)$ as an expanded network, where N represents the set of nodes, and A represents the set of arcs. Figure 2 illustrates the details of a transit stop in the expanded network. For each line that stops at a transit stop, corresponding transfer and dummy transit nodes are established to model passenger boarding and alighting behavior. Passengers can also transfer between stops using the walking arcs connecting the transfer nodes. For simplicity, we assume that demand is generated (both originated and attracted) at the stops. In this paper, N_{tf} and N_{tt} represent the sets of transfer and dummy transit nodes, and the network node set is denoted as $N = N_{tf} \cup N_{tt}$. Four different types of arcs, denoted as A_b, A_a, A_t, and A_w, constitute the expanded network, corresponding to the boarding arcs, alighting arcs, transit arcs, and walking arcs, respectively. Therefore, the set of all arcs is represented as $A = A_b \cup A_a \cup A_t \cup A_w$. Figure 3 presents a schematic diagram of the expanded network depicted in Figure 1.

A hyperpath k is defined as a subgraph $\mathcal{G}_k = (N^k, A^k, \mathbf{e}^k)$ of an expanded network $\mathcal{G}$, where the hyperpath node set $N^k \subseteq N$, the hyperpath arc set $A^k \subseteq A$, and $\mathbf{e}^k$ is the arc approaching probability vector. The hyperpath depicted in Figure 4 comprises two simple paths, denoted as k_1 and k_2, each associated with an approaching probability, denoted as $\mathbf{e}^{k_1}$ and $\mathbf{e}^{k_2}$. Assuming that the passengers and the arrival of vehicles on each line are randomly at stops, the hyperpath can provide passengers with multiple route options to reach their destinations while minimizing the expected travel cost

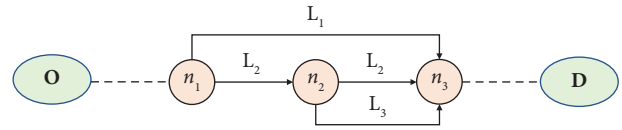

Figure 1: Illustration of the transit network.

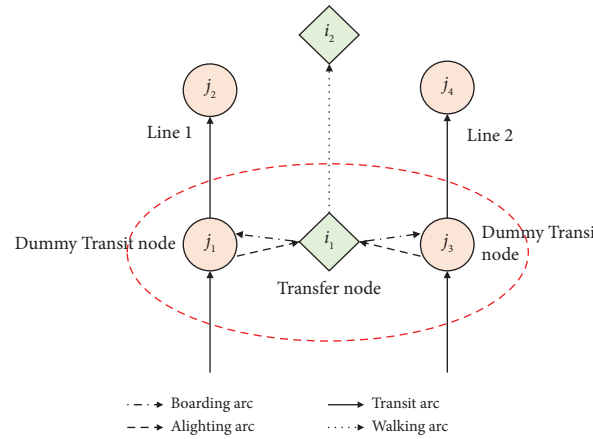

Figure 2: Illustration of a transit station in an expanded network.

using the available set of lines. For a subgraph $\mathcal{G}_k$ to qualify as a hyperpath k, it must satisfy the following four conditions: (1) $\mathcal{G}_k$ is a directed acyclic network with at least one arc; (2) for any hyperpath k, the starting node r has no predecessor, and the destination node s has no successor; (3) for any node $n \in N^k - \{r, s\}$, there exists at least one path between the starting node r and the destination node s, and if $n \notin N_{tf}$, it has at most one connected successor; and (4) the vector $\mathbf{e}^k$ satisfies the following expression:

$$e_{ij}^k > 0, \quad \forall (i,j) \in A^k, \tag{1a}$$

$$\sum_{j \in O(i)} e_{ij}^k = 1, \quad \forall i \in N^k, \tag{1b}$$

where e_{ij}^k denotes the probability that passengers use arc (i, j) at transfer node i, and $O(i)$ denotes the set of arcs starting from node i.

2.2. Transit Assignment Model

2.2.1. Arc Cost Functions considering Congestion Effects. Figure 2 illustrates four types of arcs as follows: (1) a walking arc (i_1, i_2), which connects two transfer nodes within an acceptable walking distance, depicting the walking process of the passenger between adjacent transfer nodes; (2) a transit arc (j_1, j_2), which connects two adjacent dummy transit nodes on a line, depicting the in-vehicle travel process of the passenger; (3) an alighting arc (j_1, i_1), which connects a dummy transit node to a corresponding transfer node, depicting the alighting process of the passenger; and (4) a boarding arc (i_1, j_1), which connects a transfer node to a corresponding transit node, representing the boarding process of the passenger.

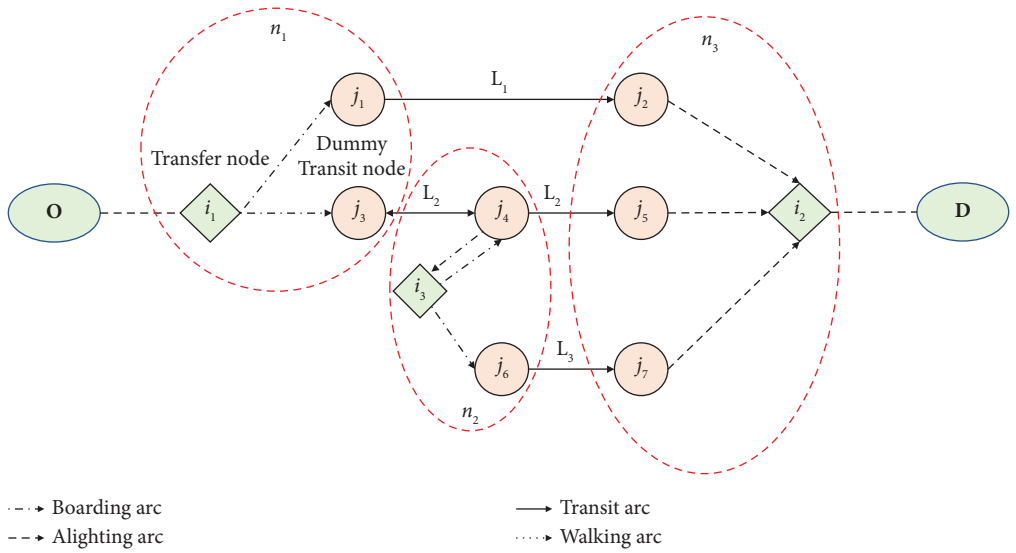

Figure 3: Illustration of an expanded network.

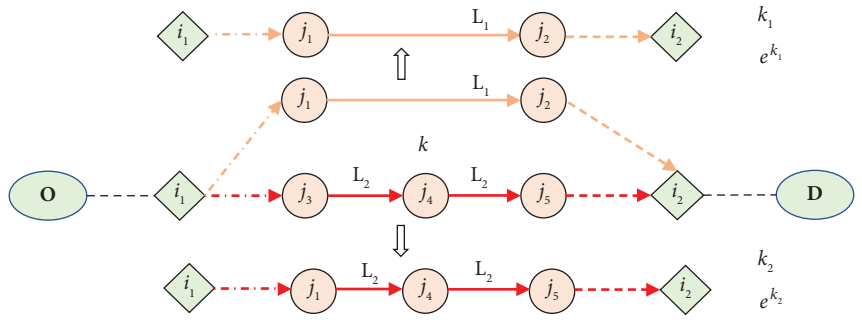

Figure 4: Illustration of a hyperpath in an expanded network.

This paper considers two types of crowding effects: boarding crowdedness and in-vehicle crowdedness, with specific definitions as outlined in Wu et al. [22], to appropriately account for the impact of crowding effects, which constitute congestion effects, on route choice. To manage the effect of congestion, travel time is generalized into travel cost on each arc, with arc cost expressed as functions of passenger flow. Corresponding to the four types of arcs illustrated in Figure 1, we define the following arc cost functions based on Wu et al. [22]:

$$\textbf{A walking arc}\,(i_1, i_2)\colon\ c_{i_1 i_2} = \alpha_1 \bar{t}_{i_1 i_2}, \quad \forall (i_1, i_2) \in A_w, \tag{2}$$

$$\textbf{A transit arc}\,(j_1, j_2)\colon\ c_{j_1 j_2} = \alpha_2 \bar{t}_{j_1 j_2} + \beta_1 \left(\frac{\left(v_{j_1 j_2} - v_{i_1 j_1}\right) + \gamma_1 v_{i_1 j_1}}{\kappa f_{i_1 j_1}} \right)^m, \quad \forall (j_1, j_2) \in A_t, \tag{3}$$

$$\textbf{An alighting arc}\,(j_1, i_1)\colon\ c_{j_1 i_1} = \alpha_3 \bar{t}_{j_1 i_1}, \quad \forall (j_1, i_1) \in A_a, \tag{4}$$

$$\textbf{A boarding arc}\,(i_1, j_1)\colon\ c_{i_1 j_1} = \beta_2 \left(\frac{v_{i_1 j_1} + \gamma_2 \left(v_{j_1 j_2} - v_{i_1 j_1}\right)}{\kappa f_{i_1 j_1}} \right)^m, \quad \forall (i_1, j_1) \in A_b, \tag{5}$$

where $\bar{t}_{i_1 i_2}$ represents the average walking time on the walking arc; $\bar{t}_{j_1 j_2}$ represents the average in-vehicle travel time on the transit arc; $\bar{t}_{j_1 i_1}$ represents the average time loss of the alighting arc; α_1, α_2, and α_3 represent the value of the time for walking, in-vehicle, and alighting, respectively; β_1, β_2, and m are all positive parameters that are typically calibrated using historical data; $v_{i_1 j_1}$ represents the boarding flow; $v_{j_1 j_2} - v_{i_1 j_1}$ represents the in-vehicle flow; γ_1 (>0) represents

the asymmetric crowding effect induced by the boarding flow on the in-vehicle flow; γ_2 (>0) represents the asymmetric crowding effect induced by the in-vehicle flow on the boarding flow; κ represents the vehicle capacity; and $f_{i_1 j_1}$ represents the corresponding departure frequency on the boarding arc.

Note that the cost of waiting arcs and transit arcs depends not only on their flow but also on the flow of adjacent arcs. Specifically, the cost of the travel arc is influenced by the average in-vehicle travel time and discomfort induced by in-vehicle crowdedness. This discomfort is associated with passengers waiting to board (boarding flow) and those already on board (in-vehicle flow) on the same line. In addition, the crowding cost on the boarding arc is determined by the boarding flow and the in-vehicle flow.

In addition to the crowding effects, passengers also experience waiting cost at each transfer node, known as the queuing effect. It is important to note that this cost is not directly associated with the arcs. Instead, a set of boarding arcs can share this cost by addressing the common line problem. As illustrated in Figure 5, multiple transit lines pass through transfer node i, denoted as $L_i = \{(i, j_1), (i, j_2), \cdots\}$, which represents the set of boarding arcs originating from the node i, where each boarding arc (i, j) corresponds to a specific line with a departure frequency of f_{ij}, measured in vehicles per hour in this paper. The route choice of the passenger at the transfer node follows a set of defined rules known as "strategies" [14]. When employing this strategy, passengers utilize the attraction set $R_i \in L_i$ to board the first arriving bus in the set and proceed towards their destinations. Assuming that (1) passengers lack real-time vehicle information, (2) they arrive at random stops without regard for established schedules, (3) they can accurately estimate the remaining travel time after boarding, and (4) the distribution of line headway is exponential [24, 25], the expected waiting time ϖ_{R_i} for passengers utilizing the attraction set R_i at the transfer node i is as follows, as stated by Li et al. [26]:

$$\varpi_{R_i} = \frac{1}{\sum_{(i,j)\in R_i} f_{ij}}. \tag{6}$$

The probability of passengers choosing to use arc (i, j), denoted as e_{ij}, can be expressed as follows:

$$e_{ij} = \frac{f_{ij}}{\sum_{(i,j')\in R_i} f_{ij'}}. \tag{7}$$

The attraction set R can be obtained using a greedy algorithm under the assumptions of the common line problem [25, 26], with given cost c_{ij} for each boarding arc, expected travel cost u^{jd} from node j to destination d, and the expected waiting time.

2.2.2. Hyperpath Cost Functions. We define the elementary path set on hyperpath k as ψ_k, and the probability of selecting an elementary path $p \in \psi_k$ is defined as π_p^k. Thus,

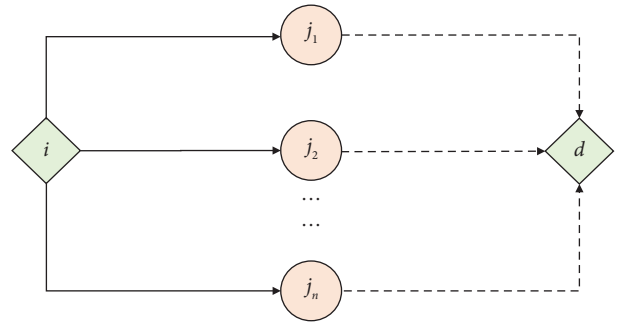

Figure 5: Illustration of the common line problem.

we obtain the following equation, according to Nguyen and Pallottino [27]:

$$\pi_p^k = \prod_{(i,j)\in p} \left(e_{ij}^k\right)^{\lambda_{ij}^p}, \quad \forall p \in \psi_k, \tag{8a}$$

$$\sum_{p\in\psi_k} \pi_p^k = 1, \tag{8b}$$

where λ_{ij}^p is a binary variable, with a value of 1 if arc (i, j) is on elementary path p, and 0 otherwise.

By enumerating the elementary paths p within ψ_k, the elementary path cost C_p is obtained by summing up the arc cost and the expected waiting time. Subsequently, it is weighted by the approaching probabilities to obtain the hyperpath cost C_k. We define ϖ_i^p as the waiting time at transfer node i on the elementary path p and calculate C_k using the following formula:

$$C_p = \sum_{(i,j)\in p} c_{ij} + \sum_{i\in p} \varpi_i^p, \quad \forall p \in \psi_k, \tag{9a}$$

$$C_k = \sum_{p\in\psi_k} \pi_p^k C_p. \tag{9b}$$

However, it is challenging to enumerate all elementary paths for large-scale transit networks. Nevertheless, it is possible to avoid enumeration by exploiting the acyclic property of hyperpaths [27]. The probability of using arc (i, j) on hyperpath k is defined as π_{ij}^k, and based on the network topology order, we have the following formula:

$$\theta_r^k = \theta_s^k = 1, \tag{10a}$$

$$\theta_j^k = \sum_{(i,j)\in I(j)} \theta_i^k e_{ij}^k, \quad \forall j \in N^k \backslash r, \tag{10b}$$

$$\pi_{ij}^k = \theta_i^k e_{ij}^k, \quad \forall (i, j) \in A^k, \tag{10c}$$

where r and s represent the starting and ending points of hyperpath k, θ_i^k denotes the probability of hyperpath k passing through node i, and $I(j)$ represents the set of arcs entering the node j.

Therefore, the formula for calculating the hyperpath cost C_k is as follows:

$$C_k = \sum_{(i,j)\in A^k} \pi_{ij}^k c_{ij}(v) + \sum_{i\in N^k \cap N_{tf}} \frac{\theta_i^k}{\sum_{(i,j')\in R_i^k} f_{ij'}}, \quad \forall \omega \in W, k \in K_w, \tag{11}$$

where R_i^k represents the attraction set at transfer node i on hyperpath k, and K_w represents the set of hyperpaths for O-D pair w. The left half of formula (11) represents the sum of expected travel cost on all arcs, while the right half represents the sum of expected waiting times at all transfer nodes.

2.2.3. Transit Equilibrium Assignment Model. The equilibrium conditions of the transit assignment model can be expressed as follows:

$$C_k \begin{cases} = u_w, & \text{if } h_k \geq 0, \\ \geq u_w, & \text{if } h_k = 0, \end{cases} \quad \forall w \in W, k \in K_w, \tag{12}$$

where h_k represents the flow on hyperpath k, and u_w represents the minimum hyperpath cost for O-D pair w.

The hyperpath flow vector $\mathbf{h}^*$ satisfies equation (12) if and only if $\mathbf{h}^*$ is a solution to the following variational inequality problem (VIP) based on hyperpaths, i.e., finding $\mathbf{h}$ such that:

$$\mathbf{C}(\mathbf{h}^*)^T(\mathbf{h} - \mathbf{h}^*) \geq 0, \quad \forall \mathbf{h} \in \Omega, \tag{13a}$$

$$\text{s.t. } \Omega \triangleq \{\mathbf{h} \mid \mathbf{q} = \wedge \mathbf{h}, \mathbf{v} = \pi \mathbf{h}, \mathbf{h} \geq 0\}, \tag{13b}$$

where $\mathbf{h}^*$ represents the vector of hyperpath flow at equilibrium, $\mathbf{h}$ represents the vector of hyperpath flow, $\mathbf{C}(\mathbf{h}^*)$ represents the cost vector of hyperpath flow at equilibrium, Ω represents the set of feasible hyperpaths, $\mathbf{q}$ represents the vector of O-D demand, $\wedge$ represents the O-D-hyperpath incidence matrix, $\mathbf{v}$ represents the vector of arc flow, and π represents the arc-hyperpath probability incidence matrix.

It is worth noting that the Jacobian matrix of the arc cost function vector $\mathbf{c}(\mathbf{v})$ exhibits asymmetry. Therefore, the assignment submodel constitutes an asymmetric transit equilibrium problem, making it challenging to express its objective function (13a) directly. However, if $\mathbf{c}(\mathbf{v})$ is strictly monotonic and continuous, i.e., the Jacobian matrix of $\mathbf{c}(\mathbf{v})$ is positive definite, then the arc flow $\mathbf{v}$ at equilibrium is also unique [21]. Further discussion of the mathematical properties of the model can be found in the literature [21, 22].

3. Frequency Optimization Model and Solution Algorithm

3.1. Bilevel Programming Model Formulation. In general, the definition of a bilevel programming problem is as follows:

$$(U) \min_{\mathbf{x}} \ F(\mathbf{x}, \mathbf{y}(\mathbf{x})), \tag{14a}$$

$$\text{s.t.} \quad \mathbf{G}(\mathbf{x}, \mathbf{y}(\mathbf{x})) \leq 0, \tag{14b}$$

where $\mathbf{y} = \mathbf{y}(\mathbf{x})$ is obtained by solving the lower-level optimization model:

$$(L) \min_{\mathbf{y}} \ f(\mathbf{x}, \mathbf{y}), \tag{14c}$$

$$\text{s.t.} \quad \mathbf{g}(\mathbf{x}, \mathbf{y}) \leq 0. \tag{14d}$$

In the above bilevel programming model, U represents the upper-level model, where $F(\mathbf{x}, \mathbf{y}(\mathbf{x}))$ is the objective function of the upper-level decision-makers, $\mathbf{x}$ is the decision variables of the upper-level decision-makers, and $\mathbf{G}(\mathbf{x}, \mathbf{y}(\mathbf{x}))$ represents the constraint set for the upper-level decision variables. L represents the lower-level model, where $f(\mathbf{x}, \mathbf{y})$ is the objective function of the lower-level decision-makers, $\mathbf{y}$ is the decision variables of the lower-level decision-makers, and $\mathbf{g}(\mathbf{x}, \mathbf{y})$ represents the constraint set for the lower-level decision variables. It is worth noting that the lower-level decision variables $\mathbf{y}$ can be expressed as a function of the upper-level decision variables $\mathbf{x}$, i.e., $\mathbf{y} = \mathbf{y}(\mathbf{x})$, and are often referred to as the response functions. In the upper-level model, its objective function depends not only on its own decision variables but also on the optimal solution of the lower-level model. Conversely, the optimal solution of the lower-level model is influenced by the decision variables of the upper-level model. This implies that the solution of the bilevel programming model is achieved when the entire decision system reaches an equilibrium state, where the upper-level decision-makers achieve the optimal objective, and the lower-level decision-makers achieve the optimal state within the constraints.

Bilevel programming models are widely utilized in the transportation field, especially for establishing toll standards and designing transportation networks. In general, transportation network design problems involve two main stakeholders: transportation operators and passengers. Passengers aim to achieve the best travel experience, which may lead to increased operational cost. Transportation operators, on the other hand, seek to minimize the overall travel cost of the transit system while controlling operational expenses. The interests of transportation operators and passengers are interconnected yet conflicting. To reconcile these competing interests, bilevel programming models are a viable solution.

In this paper, we present a bilevel programming model for frequency optimization based on the characteristics of the transit network. The upper-level model seeks to optimize the travel cost of the transit system while controlling the operational cost. The lower-level model is to accurately estimate the impact of frequency changes on passenger travel behavior.

$$(U_1) \min_{\mathbf{v}} \ Z(\mathbf{f}, \mathbf{v}(\mathbf{f})) = \rho_1 \left(\sum_{s\in A_s} c_s[v_s(\mathbf{f})] v_s(\mathbf{f}) \right) + \rho_2 \left(\theta \sum_{l\in L} F_l f_l \right), \tag{15a}$$

$$\text{s.t. } f_l \geq 1, \quad \forall l \in L, \tag{15b}$$

where $v_s(\mathbf{f})$ represents the arc flow determined by the lower-level model L_1:

$$(L_1)\ \mathbf{C}(\mathbf{h}^*)^T(\mathbf{h}-\mathbf{h}^*)\geq 0,\quad \forall \mathbf{h}\in\Omega, \tag{15c}$$

$$\text{s.t. } \Omega \triangleq \{\mathbf{h}\mid \mathbf{q}=\wedge\mathbf{h}, \mathbf{v}=\pi\mathbf{h}, \mathbf{h}\geq 0\}, \tag{15d}$$

where ρ_1 denotes the weight coefficient for the travel cost of the transit system, c_s denotes the cost of arc s, v_s denotes the flow on arc s, ρ_2 denotes the weight coefficient for the operational cost of the transit system, θ denotes the coefficient for converting the operational cost at the current frequency into the travel cost, F_l denotes the operational cost per unit frequency of the transit line, f_l denotes the frequency of the transit line, L denotes the set of transit lines within the transit network, and the definition of Ω can be referenced in the definitions of formulas (13a) and (13b).

The upper-level model, denoted as U_1, is a continuous network design model with the objective function aimed at minimizing the weighted sum of the transit system's travel cost and operating cost. Constraints (15b) represent the constraints of the upper-level model, ensuring that at least one bus departs from each transit route per unit time. The lower-level model, referred to as L_1, is a hyperpath-based transit equilibrium assignment model. Equation (15d) constitutes the constraints of the lower-level model, encompassing flow conservation and non-negativity constraints. As depicted in Figure 6, given a set of decision variables $\mathbf{f}$, the upper-level decision-makers adjust their decisions based on the equilibrium flow distribution derived from the lower-level model, resetting frequencies to minimize the upper-level objective function. This process is iterated until the bilevel programming model reaches an equilibrium state.

3.2. Solution Algorithm. Based on formulations (15a) and (15b), the upper-level model U_1 is a constrained nonlinear optimization problem concerning departure frequency, which can be solved using established methods such as the outer approximation method and the penalty function method. However, these approaches involve dealing with penalty functions or multipliers, making the computation relatively difficult. Therefore, we propose a heuristic gradient descent algorithm to solve this problem, which computes the gradient of the objective function at the current point, searches for the stepsize within the feasible region, and iteratively updates until a convergence criterion is met. The algorithm framework is described in Algorithm 1.

Since we use a local search strategy, the solution will depend on the initial point during the iteration process. Therefore, the historical departure frequencies for transit lines are naturally chosen as the initial solution in step 1 of the algorithm.

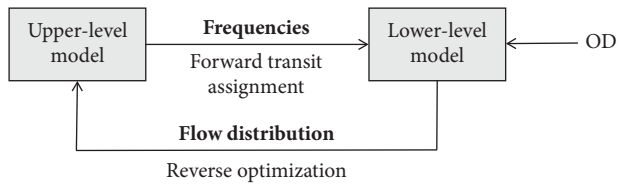

Figure 6: Illustration of the bilevel programming mode.

Step 3 will be discussed in Section 3.3 since the calculation of the descent direction involves determining the rate of change of the arc flow $\mathbf{v}$ under perturbations of the departure frequencies $\mathbf{f}$.

There are several approaches for choosing the convergence criteria for the algorithm in Step 4. The norm of the search direction, the (relative) improvement of the objective function value in the last iterations, the maximum number of iterations, the maximum runtime, or any combination of these factors are all viable options. Since the bilevel programming model represents a nonconvex optimization problem, finding an optimal solution for the model remains an NP-hard problem [28]. Therefore, the chosen convergence criteria should guarantee that the algorithm converges to a reasonably good local optimum. In this paper, an iteration with a relatively small improvement in the objective function value is defined as bad_iter, and the successive occurrence of bad_iter is counted as bad_iter_num. If bad_iter_num is relatively large, the algorithm is considered to have converged to a local optimum. However, if bad_iter_num exceeds a certain threshold, it can be concluded that the algorithm cannot converge to a better local optimum within a reasonable amount of time. Therefore, a combination of the relative improvement in the objective function value, the maximum number of bad_iter, and the maximum number of iterations is chosen as the convergence criteria for the algorithm.

Because $\mathbf{v}$ is a nonlinear implicit function of $\mathbf{f}$, an accurate linear search in Step 5 requires a high computational cost. Therefore, the nonexact step search method called the Armijo strategy [29] is employed in this study. For each trial stepsize, we performed a transit equilibrium assignment to obtain the objective function value. The search was terminated if the improvement of the objective function value exceeded a certain threshold; otherwise, it was reset and the preceding steps were repeated. In practice, this search method may involve a large number of transit equilibrium assignments, imposing a significant computational burden. Nonetheless, the advanced GP-AIL algorithm employed in this paper can solve the transit equilibrium assignment problem accurately in a short time, making it suitable for large-scale networks [21]. Figure 7 depicts the flowchart of Algorithm 1.

Step 1: Initialization: Determine the initial departure frequencies $\mathbf{f}^0$, the maximum number of iterations $n_{\max}$, the maximum number of bad iterations max_bad_iter, the convergence accuracy ε_1, and the absolute improvement in the objective function value ε_2; set the number of iterations $n = 0$ and bad_iter_num = 0.
Step 2: Solve the lower-level problem: Use the GP-AIL algorithm [21] to solve the transit equilibrium assignment problem to obtain the equilibrium flow solution $\mathbf{v}^{*n}$ and calculate the upper-level objective function value Z^n.
Step 3: Compute the descent direction: Calculate an approximate gradient $\nabla Z(\mathbf{f}^n)$ of the objective function, and then the descent direction is determined as $\mathbf{r}^n = -\nabla Z(\mathbf{f}^n)$.
Step 4: Check the convergence condition:
Step 4.1 Determine whether the convergence condition $|(Z^n - Z^{n-1})/Z^{n-1}| < \varepsilon_1$ is satisfied; if yes, bad_iter_num+ = 1; if no, bad_iter_num = 0.
Step 4.2 Determine whether the convergence condition $|(Z^n - Z^{n-1})/Z^{n-1}| < \varepsilon_1$ and bad_iter_num > max_bad_iter or $n > n_{\max}$ is satisfied; if so, the algorithm terminates; if not, go to Step 5.
Step 5: Determine the stepsize:
Step 5.1 Determine the maximum step $\alpha^n_{\max}$, so that the updated frequencies satisfy the constraint (15b). $\alpha^n_{\max} = \min\{+\infty, -(f^n_l - 1)/r^n_l : r^n_l < 0, \forall l \in L\}$.
Step 5.2 Search for the optimal step α^n:
Step 5.2.1 Set $\alpha^0 = \alpha^n_{\max}$, the search iteration number $j = 0$, the maximum search iteration number $j_{\max}$, and $Z^{(n,0)} = Z(\mathbf{f}^n)$.
Step 5.2.2 Obtain the updated equilibrium flow solution $\mathbf{v}^{*j}$ pertaining to the updated frequencies by employing transit equilibrium assignment, and calculate the corresponding value of the upper-level objective function: $Z^{(n,j)} = Z(\mathbf{f}^n + \alpha^j \mathbf{r}^n)$.
Step 5.2.3 If the convergence condition $Z^{(n,0)} - Z^{(n,j)} > \varepsilon_2$ is satisfied, set $\alpha^n = \alpha^j$ and stop the algorithm; otherwise, proceed to Step 5.2.4.
Step 5.2.4 If j is equal to $j_{\max}$, set $\alpha^n = \text{argmin}_{\alpha^j}(Z^{(n,j)})$; otherwise, set $\alpha^j = \alpha^j/\theta$, $j = j + 1$, and go to Step 5.2.2.
Step 6: Update: Set $\mathbf{f}^n = \mathbf{f}^n + \alpha^n \mathbf{r}^n$, $n = n + 1$ and go to Step 2.

Algorithm 1: The heuristic gradient algorithm framework for frequency optimization.

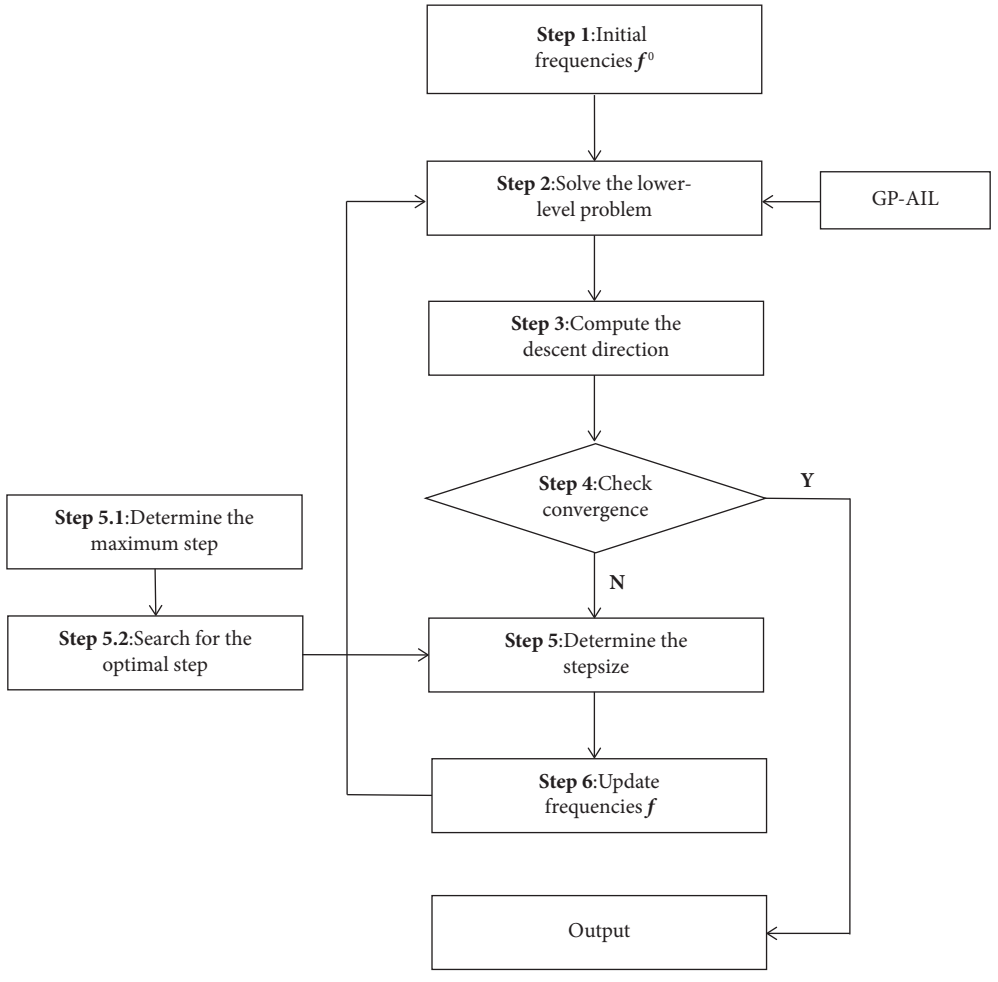

Figure 7: Flowchart of Algorithm 1.

3.3. Approximation of the Gradient. In order to construct the descent direction for the upper-level model U_1, we compute the gradient of the objective function Z with respect to $\mathbf{f}$ using the following formula:

$$F_1\left(v_s(\mathbf{f})\right) = c_s\left(v_s(\mathbf{f})\right)v_s(\mathbf{f}), F_2(\mathbf{f}) = \theta\sum_{l\in L}F_l f_l, \tag{16a}$$

$$Z(\mathbf{f}) = \rho_1\left(\sum_{s\in A_s}F_1\left(v_s(\mathbf{f})\right)\right) + \rho_2 F_2(\mathbf{f}), \tag{16b}$$

$$\nabla Z(\mathbf{f}) = \rho_1\left(\sum_{s\in A_s}\nabla_{\mathbf{f}}v_s(\mathbf{f})\nabla_{\mathbf{v}}F_1\left(v_s(\mathbf{f})\right)\right) + \rho_2\nabla_{\mathbf{f}}F_2(\mathbf{f}), \tag{16c}$$

$$\nabla_{\mathbf{f}}v_s(\mathbf{f}) = \left\{\frac{\partial v_s(\mathbf{f})}{\partial f_l}\right\}, \quad \forall l \in L, \tag{16d}$$

$$\nabla_{\mathbf{v}}F_1\left(v_s(\mathbf{f})\right) = \left\{\frac{\partial c_s\left[v_s(\mathbf{f})\right]}{\partial v_s}v_s(\mathbf{f}) + c_s\left[v_s(\mathbf{f})\right]\right\}, \quad \forall l \in L, \tag{16e}$$

$$\nabla_{\mathbf{f}}F_2(\mathbf{f}) = \{\theta F_l\}, \quad \forall l \in L. \tag{16f}$$

It is noteworthy that both partial derivatives $\nabla_{\mathbf{v}}F_1\left(v_s(\mathbf{f})\right)$ and $\nabla_{\mathbf{f}}F_2(\mathbf{f})$ have relatively straightforward analytical expressions for calculation. However, owing to the implicit definition of the relationship between arc flow $\mathbf{v}$ and departure frequencies $\mathbf{f}$, direct representation through analytical expressions is not feasible. Therefore, it is challenging to calculate the gradient for the nonlinear implicit function $v_s(\mathbf{f})$, which will be calculated in this paper using sensitivity analysis for transit equilibrium problems.

The restriction approach, first introduced by Tobin and Friesz [30] and later extended by Yang and Bell [31], is employed to create a constrained equation system that meets the requirements for applying the sensitivity analysis method in nonlinear programming. Building upon this research, Du and Chen [32] derived the expressions for sensitivity analysis based on the variational inequality (VI) form of transit equilibrium assignment models. By applying the Karush–Tucker (K-T) conditions to the optimal solution $\mathbf{h}^*$ with the perturbations $\epsilon = 0$, we obtain the following equations:

$$\mathbf{C}\left(\mathbf{h}^*, 0\right) - \boldsymbol{\lambda} - \wedge^T\boldsymbol{\mu} = 0, \tag{17a}$$

$$\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{h}^* = 0, \tag{17b}$$

$$\wedge\mathbf{h}^* - \mathbf{q}(0) = 0, \tag{17c}$$

$$\boldsymbol{\lambda} \geq 0, \tag{17d}$$

$$\mathbf{h}^* \geq 0, \tag{17e}$$

where $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ represent the Lagrange multiplier vectors corresponding to $\mathbf{h}^* \geq 0$ and $\wedge\mathbf{h}^* - \mathbf{q}(0) = 0$, respectively.

Under Assumptions 1 and 2, Du and Chen [32] proposed that the equilibrium solution is differentiable at the current point.

Assumption 1. The cost functions $\mathbf{c}(\mathbf{v})$ of boarding arcs and transit arcs are increasing functions of arc flow $\mathbf{v}$, and the travel cost depends primarily on their own flow, meaning the derivative of $\mathbf{c}(\mathbf{v})$ with respect to their own flow is greater than the derivative with respect to the flow on related arcs.

Assumption 2. The path flow solution $\mathbf{h}^*$ at the equilibrium point is nondegenerate.

Assumption 1 ensures that $\mathbf{c}(\mathbf{v})$ is once continuously differentiable concerning arc flow $\mathbf{v}$ and perturbation vector ϵ. Assumption 2 excludes the existence of degenerate equilibrium path solutions and ensures the differentiability of the equilibrium solution at the current point. Note that preserving the linearly independent fraction of all equilibrium hyperpaths is equivalent to selecting a maximum set of linearly independent columns of the incidence matrix $\left[\wedge^T\ \pi^T\right]^T$. The linearly independent equilibrium hyperpaths to be retained, denoted as $\mathbf{h}^0$, correspond to this set of column vectors. Using this approach, we constrain the equilibrium network while preserving its independent hyperpaths. Formulas (17a)–(17e) can then be simplified into the following system of equations:

$$\mathbf{C}^0(\mathbf{h}^*, 0) - \wedge^{0T}\boldsymbol{\mu} = 0, \tag{18a}$$

$$\wedge^0 \mathbf{h}^{0*} - \mathbf{q}(0) = 0, \tag{18b}$$

where $\mathbf{C}^0$ represents the path cost vector corresponding to $\mathbf{h}^0$, and $\wedge^0$ represents the incidence matrix corresponding to $\mathbf{h}^0$.

As a result, by calculating the Jacobian matrices of the system (18a) and (18b) with respect to $(\mathbf{h}, \boldsymbol{\mu})$, and ϵ, we can obtain the following formulas:

$$\mathbf{J}_{\mathbf{h},\boldsymbol{\mu}}(0) = \begin{bmatrix} \nabla_{\mathbf{h}}\mathbf{C}^0(\mathbf{h}^*, 0) & -\Lambda^{0T} \\ \Lambda^0 & 0 \end{bmatrix}, \tag{19a}$$

$$\mathbf{J}_{\epsilon}(0) = \begin{bmatrix} \nabla_{\epsilon}\mathbf{C}^0(\mathbf{h}^*, 0) \\ -\nabla_{\epsilon}\mathbf{q}(0) \end{bmatrix}, \tag{19b}$$

$$\begin{bmatrix} \nabla_{\epsilon}\mathbf{h}^0 \\ \nabla_{\epsilon}\boldsymbol{\mu}^0 \end{bmatrix} = \left[\mathbf{J}_{\mathbf{h},\boldsymbol{\mu}}(0)\right]^{-1}\left[-\mathbf{J}_{\epsilon}(0)\right]. \tag{19c}$$

Du and Anthony also proved that the Jacobian matrix $\mathbf{J}_{\mathbf{h},\boldsymbol{\mu}}(0)$ is invertible with the linear independence of the equilibrium hyperpath set associated with the incidence matrix $\begin{bmatrix} \wedge^{0T} & \pi^{0T} \end{bmatrix}^T$, and the result of $\nabla_{\epsilon}\mathbf{v}$ is independent of the choice of the hyperpaths.

The perturbation vector $\mathbf{f}$ in the frequency optimization problem represents frequency variations. The following is the formula for determining the derivative of arc flow $\mathbf{v}$ with respect to$\mathbf{f}$:

$$\mathbf{J}_{\mathbf{f}}(0) = \begin{bmatrix} \nabla_{\mathbf{f}}\mathbf{C}^0(\mathbf{h}^*, 0) \\ 0 \end{bmatrix}, \tag{20a}$$

$$\begin{bmatrix} \nabla_{\mathbf{f}}\mathbf{h}^0 \\ \nabla_{\mathbf{f}}\boldsymbol{\mu}^0 \end{bmatrix} = \left[\mathbf{J}_{\mathbf{h},\boldsymbol{\mu}}(0)\right]^{-1}\left[-\mathbf{J}_{\mathbf{f}}(0)\right], \tag{20b}$$

$$\nabla_{\mathbf{f}}\mathbf{v} = \pi'\mathbf{h}^0 + \pi\nabla_{\mathbf{f}}\mathbf{h}^0, \tag{20c}$$

where π' is composed of the derivatives of π_s^k which denotes the probability that hyperpath k uses arc s with respect to $\mathbf{f}$.

Therefore, by substituting equation (20c) into equation (16c), the approximate gradient of the objective function Z with respect to $\mathbf{f}$ can be obtained.

4. Numerical Results

This section is organized into three main parts. The first part provides a detailed description of the algorithm's computational environment, the test network, and specific implementation details. The second part examines the convergence characteristics of the test algorithm and compares the application performance of the algorithm before and after considering congestion effects on a medium-sized network. The last part further examines the computational results of the algorithm in a large-scale network.

4.1. Computing Environment and Algorithm Implementation Details. The algorithms tested in this section were compiled using the Toolkit of Network Modeling, a C++ class library specialized in modelling transportation networks [33]. All numerical results reported in this section were obtained on a Windows 10 64 bit PC with Intel® Core™ i7-11700 CPU 2.50 GHz and 64 G RAM. The performance of the test algorithms is evaluated by applying them to solving the Sioux-Falls and Winnipeg networks. The Sioux-Falls network, a medium-scale network, was provided by Szeto and Jiang [34], as shown in Figure 8, while the Winnipeg network is a real large-scale network from the city of Winnipeg, with an Emme3 demo project (See https://www.inrosoftware.com/en/products/emme/), as shown in Figure 9. The topology of each network is described in Table 1.

Table 2 presents the parameter values used in the arc cost functions, which are adopted from Wu et al. [22]. Table 3 presents the values of the basic parameters used in the algorithm. The coefficient θ, which represents the conversion of operating cost under the current departure frequency into travel cost, is set to 1/30 (hours per yuan). According to the transit operational cost estimation index system introduced by Huanghn [35], the operational cost mainly includes labor cost, vehicle cost, station cost, fuel cost, etc. The value of F_l specified in this paper amounts to 1,000 yuan per unit departure frequency, which represents the operational cost incurred for each unit increase in the frequency of the transit line l. The weight ratio ρ_1: ρ_2 between travel cost and operational cost in the transportation system is set to be 60: 1. This configuration aims to increase the attractiveness of the transit system by increasing the operating cost to some extent while reducing the travel cost of passengers. It is worth noting that this weight ratio can be flexibly adjusted according to specific circumstances. The values of $j_{\max}$, $n_{\max}$, and max_bad_iter are set based on empirical knowledge from numerical experiments and can be flexibly adjusted according to actual circumstances. The values of ε_1 and ε_2 depend on the input of specific problems and can be adjusted according to actual situations.

A few implementation details of the algorithm are given below. (1) We employ the relative gap (RG) as the convergence criterion for the transit assignment model. The formula for calculating RG is as follows:

$$\mathrm{RG} = 1 - \frac{\sum_{w\in w} u_w q_w}{\sum_{s\in A} c_s v_s + \tau}, \tag{21}$$

where τ denotes the total waiting cost, $\sum_{s\in A} c_s v_s$ denotes the total travel cost, and u_w denotes the minimum travel cost for OD pair w. Note that the precision of RG affects the accuracy of the passenger distribution prediction. In this section, we set the RG precision to 10^{-7} for the medium-sized network and 10^{-4} for the large-scale network, because this level of precision satisfies practical requirements. (2) The method in the lower-level transit equilibrium model that does not consider congestion effects ignores both queuing and crowding effects. Specifically, the expected waiting time for passengers at transfer nodes, denoted as $\bar{\omega}_{R_i}$, is set to 0, and the boarding and in-vehicle crowding functions for boarding

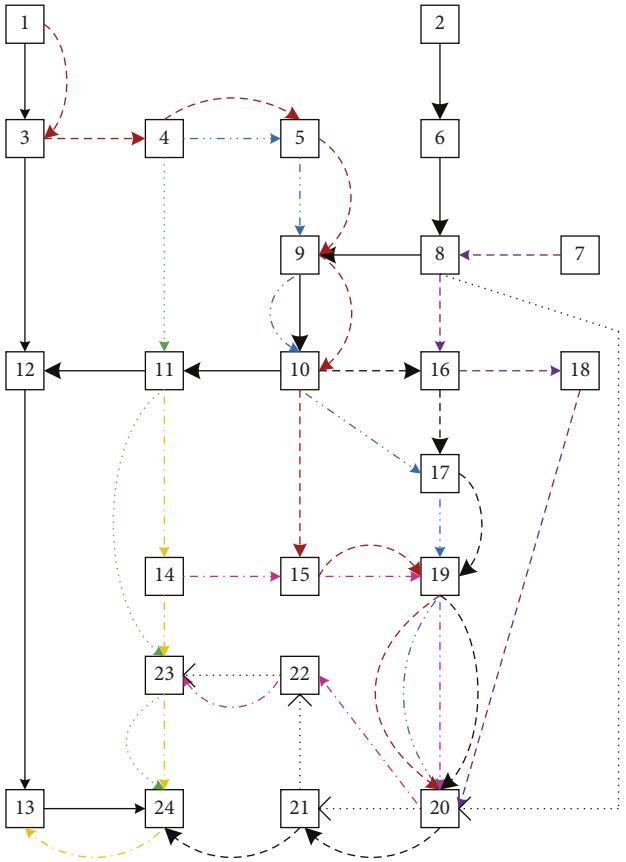

Figure 8: Topology of the Sioux-Falls network.

Figure 9: Topology of the Winnipeg network.

and transit arcs are both set to 0, resulting in the following arc cost functions:

$$c_{i_1 j_1} = 0, c_{j_1 j_2} = \alpha_2 \bar{t}_{j_1 j_2}. \quad (22)$$

(3) The upper limit for the frequency is set at 30 buses per hour, i.e., $\mathbf{f}^k = \mathbf{f}^k + \alpha^k \mathbf{r}^k < 30$, with the flexibility to modify it based on specific circumstances. (4) We consider two types of congestion effects: crowding effect and queuing effect. In addition, there are two other types of effects: capacity effect and bunching effect, which are not considered in this paper.

4.2. Sioux-Falls Network Analysis. The computational results of the optimization algorithm considering congestion effects in the Sioux-Falls network (depicted in Figure 8) are shown in Figure 10, with basic network information available in Table 4. In this plot, the horizontal axis represents the computational time of the algorithm in seconds, while the vertical axis quantifies the objective function value, which reflects the weighted combination of travel cost and operational cost within the transit system, measured in hours. In general, the algorithm demonstrates rapid convergence to an optimal solution within a relatively short timeframe. The objective function value reaches convergence within the first 20 seconds. In particular, within the first ten iterations, the objective function value experiences a rapid decrease and approaches a local optimum, with a difference of no more than 0.011%. Subsequently, the algorithm continues to search for even more optimal local solutions until it reaches the maximum convergence iterations. This observation highlights the algorithm's ability to significantly improve system performance in a short timeframe, making it particularly valuable for real-world engineering applications.

We provide a frequency adjustment scheme that does not consider congestion effects and simulate both schemes to validate the rationality of the optimization scheme. Table 5 presents two departure frequency schemes and their respective travel cost (including congestion cost). Overall, the optimization scheme that considers congestion effects demonstrates superior performance in optimizing the transit system. Specifically, this optimization scheme significantly reduces the travel cost of the transit system by 3.9% compared to the initial scheme. In contrast, the adjustment scheme that does not consider congestion effects fails to reduce travel cost and actually increases them by 6.4%. This is because the lower-level transit assignment model does not account for congestion effects, fails to accurately reflect passenger travel behavior, and cannot predict the impact of changes in departure frequency on passenger distribution accurately, resulting in the inferior performance of the adjustment plan. Therefore, the optimization algorithm that considers congestion effects is more suitable for optimizing departure frequencies in transit networks.

We further examine the convergence characteristics of the algorithm. Note that the algorithm may converge to a local optimum rather than a global optimum due to the nonlinear nature of the optimization model and the heuristic gradient descent algorithm used. We compare the convergence solutions under different initial departure frequencies to determine the degree of variation. It is assumed that the algorithm can converge to a nearby local optimum if the initial frequencies are sufficiently close.

Table 6 provides four initial departure frequencies, while Figure 11 and Table 7 illustrate the corresponding optimized solutions and their respective objective function values. In general, the optimized solutions show slight fluctuations near a certain solution despite variations in initial departure frequencies. Specifically, the variations in objective function values among the optimal solutions do not exceed 0.05%. This implies that the optimization algorithm can converge to

Table 1: Basic test network information.

Networks	Size	Stops	Lines	Nodes	Arcs	O-D pairs	Trips
Sioux-falls	Medium	24	10	84	150	16	7,200
Winnipeg	Large	690	132	5,186	13,276	5,332	15,729

Table 2: Basic parameters used in the arc cost functions.

Parameter	Value
α_1	1.0
α_2	2.5
α_3	1.0
β_1	5.0
m	2
β_2	4.0
κ	30
γ_1	1.2
γ_2	0.3

Table 3: Basic parameters used in the algorithm.

Parameter	Value
θ	0.033
F_l	1.0
$j_{\max}$	2.5
$n_{\max}$	30
max _bad_iter	1,000
ρ_1	60
ρ_2	1
ε_1	10^{-5}
ε_2	1000

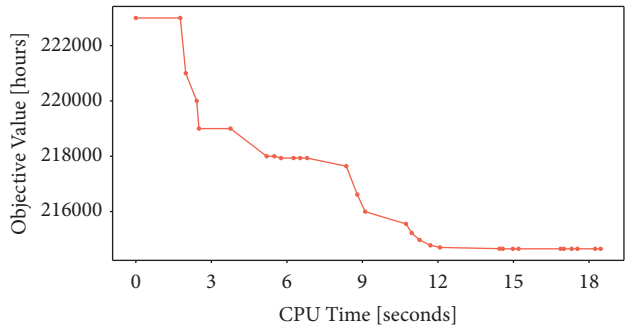

Figure 10: Convergence curve of the Sioux-Falls network objective function.

nearby local optima under different initial frequencies, further confirming the reliability of the algorithm in the context of frequency optimization.

4.3. Winnipeg Network Analysis. The computational results of the optimization algorithm for the Winnipeg network (Figure 9) are depicted in Figure 12. Overall, the algorithm demonstrates the ability to converge to an optimized solution within an acceptable amount of time, with the objective function value reaching convergence within 3.8 hours. In particular, the optimization algorithm significantly improves system performance in the first four

Table 4: Basic information of the Sioux-Falls network.

Line ID	Initial frequency	Stops
1	12	(4, 11, 23, 24)
2	13	(1, 3, 12, 13, 24)
3	10	(11, 14, 23, 24, 13)
4	12	(8, 20, 21, 22, 23)
5	10	(7, 8, 16, 18, 20)
6	14	(14, 15, 19, 20, 22, 23)
7	20	(2, 6, 8, 9, 10, 11, 12)
8	22	(4, 5, 9, 10, 17, 19, 20)
9	24	(10, 16, 17, 19, 20, 21, 24)
10	20	(1, 3, 4, 5, 9, 10, 15, 19, 20)

Table 5: Comparison of departure frequency schemes.

Congestion considered	Frequency adjustment scheme	Travel cost
Yes	(18.5, 23.0, 2.5, 2.8, 22.4, 3.2, 24.2, 19.7, 24.8, 24.1)	2,09,148
No	(8.2, 11.3, 6.9, 10.6, 6.9, 9.5, 18.0, 21.3, 16.1, 13.4)	2,31,592
Initial scheme	(12, 13, 10, 12, 10, 14, 20, 22, 24, 20)	2,17,730

Table 6: Different initial departure frequencies.

Initial scheme ID	Frequency scheme
Initial scheme 1	(12, 13, 10, 12, 10, 14, 20, 22, 24, 20)
Initial scheme 2	(13, 14, 10, 12, 10, 14, 20, 22, 24, 20)
Initial scheme 3	(12, 13, 10, 12, 10, 13, 19, 22, 24, 20)
Initial scheme 4	(12, 13, 12, 12, 10, 14, 21, 22, 23, 20)

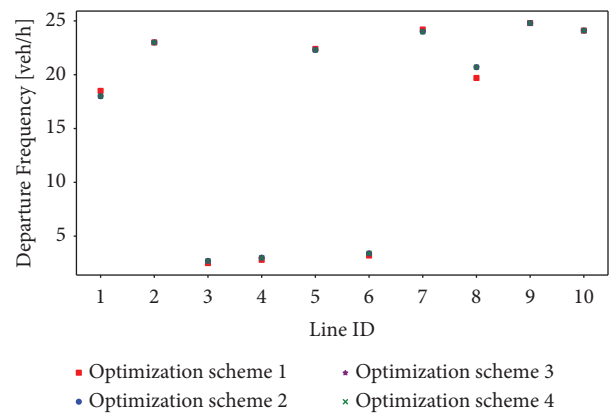

Figure 11: Schematic diagram of departure frequencies.

iterations, with the objective function value in the fourth iteration deviating from the local optimum by no more than 0.001%. The consistency of these results with the previous findings indicates that the algorithm proposed in this paper is capable of efficiently optimizing departure frequencies in

Table 7: Comparison of optimal frequencies under different initial departure frequencies.

Initial scheme ID	Frequency adjustment scheme	Objective values
Initial scheme 1	(18.5, 23.0, 2.5, 2.8, 22.4, 3.2, 24.2, 19.7, 24.8, 24.1)	2,14,652
Initial scheme 2	(18.0, 23.0, 2.7, 3.0, 22.3, 3.4, 24.0, 20.7, 24.8, 24.1)	2,14,557
Initial scheme 3	(18.1, 23.0, 2.6, 2.9, 22.4, 3.1, 24.0, 19.8, 24.8, 24.1)	2,14,662
Initial scheme 4	(19.0, 23.0, 2.9, 2.9, 22.4, 3.3, 24.3, 19.8, 24.7, 24.1)	2,14,654

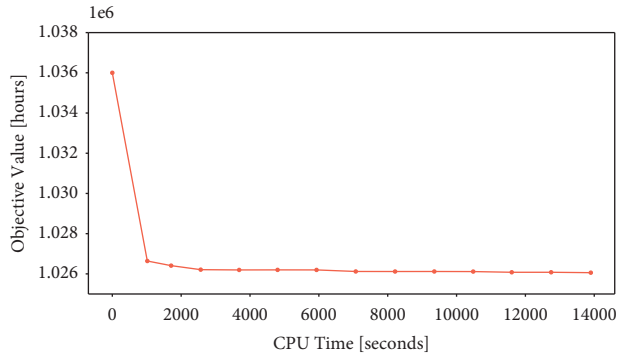

Figure 12: Convergence curve of the Winnipeg network objective function.

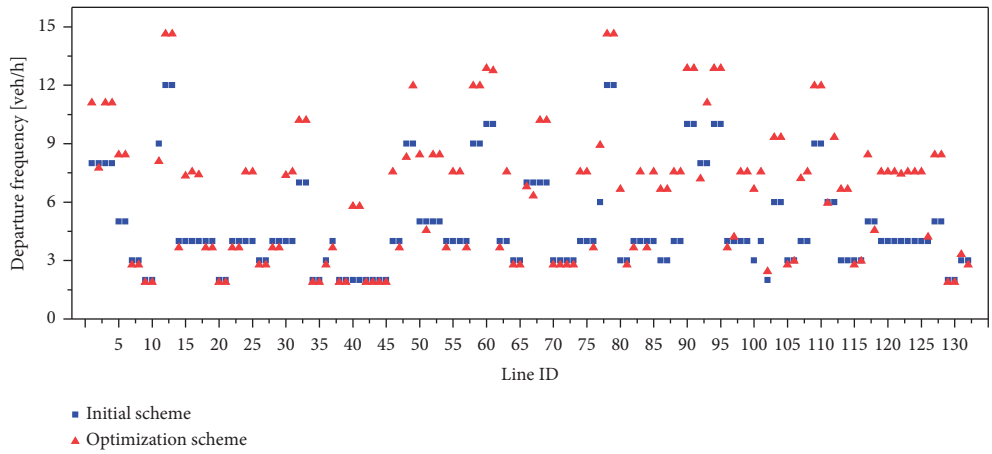

Figure 13: Comparison of departure frequency of each line before and after optimization.

both medium-scale and large-scale networks within an acceptable timeframe.

The comparison of frequencies before and after optimization is shown in Figure 13. Before optimization, the average frequency across the entire network was 4.81 buses per hour, which increased to 6.52 buses per hour after optimization. It is worth noting that there was little difference in departure frequencies between several transit lines in the Winnipeg network before and after optimization. Table 8 provides details on three bus routes (with line IDs 11, 48, 67, and 92) that experienced minimal changes in frequency, with reductions of 0.92, 0.71, 0.69, and 0.80, respectively, all of which are less than 1. This is due to the fact that the Winnipeg network used in this study is a real transit network with departure frequencies determined through extensive surveys and research. In addition, these few lines serve a limited number of stops, have shorter travel times, and experience relatively low travel demand between their corresponding origin-destination pairs. Therefore, intuitively, there is little need for significant adjustments to their departure frequencies, resulting in minimal differences in frequency before and after optimization for certain lines.

Table 8: Basic information of transit lines with minimal frequency changes.

Line ID	11	48	67	92
Operational line	Line 15	Line 32	Line 40	Line 52
Stops number	38	24	24	30
Running time (min)	72	82	66	89

Some lines show significant changes in departure frequencies before and after optimization. Table 9 presents the basic information for four transit lines that had significant frequency increases after optimization, with line IDs 24, 74, 83, and 100 exhibiting frequency increases of 3.55, 3.55, 3.55, and

Table 9: Basic information of transit lines with large frequency changes.

Line ID	24	74	83	100
Operational line	Line 21	Line 44	Line 48	Line 56
Stops number	58	69	47	41
Running time (min)	117	215	98	60

3.66, respectively. The significant differences in frequency before and after optimization are primarily related to the design and layout of these lines. These lines share common characteristics, such as passing through a higher number of stops, longer travel times, significant travel demand between their respective origin-destination pairs, and a lack of alternative transfer lines. In addition, these lines are strategically aligned with the primary flow of passengers, facilitate transfers to several other routes, and provide essential transportation options. As a result, increasing the departure frequencies of these lines can alleviate the burden on the remaining transit lines, resulting in more efficient passenger movement and ultimately optimizing system performance.

In conclusion, the optimization algorithm that considers congestion effects can efficiently optimize bus departure frequencies in a short timeframe and is applicable to large-scale transit networks.

5. Conclusion

Traditional transit frequency setting usually deviates from real-world situations due to unrealistic assumptions about passenger travel behavior, resulting in practical limitations and design constraints. This paper introduces a bilevel programming model that accurately captures the interactions between transit operators and passengers, providing reliable decision support for transit planning and policymakers. Compared to existing bilevel models, it has two advantages: first, it explicitly considers passenger congestion cost in the upper-level model, rather than solely travel time. Second, it incorporates the influence of two types of congestion effects on passenger travel behavior in the lower-level model, which realistically reflects the changes in passenger travel patterns under different frequencies.

The optimization model can effectively improve the transportation system and enhance user satisfaction by incorporating congestion effects. This is achieved by incorporating congestion effects into the lower-level transit assignment model, which better reflects the travel characteristics of transit users. Numerical results demonstrate that congestion-aware optimization significantly improves system performance, reducing travel cost (including congestion cost) by 9.7% compared to adjustments without congestion consideration.

Furthermore, the heuristic gradient descent algorithm proves to be an effective solution for departure frequency optimization. Numerical experiments show that the optimization algorithm converges to solutions in a relatively short timeframe. For the Sioux-Falls network, the algorithm reaches convergence within 20 seconds, while for the Winnipeg network, it converges within 3.8 hours.

Finally, although this paper has demonstrated the potential of the proposed algorithm to address the optimization of bus departure frequencies, it still has the following limitations: (1) the upper-level model does not consider the impact of factors such as fleet size and vehicle acquisition cost on the overall system cost; (2) the lower-level model does not consider the influence of capacity and bunching effects on user travel behavior; and (3) the lack of comparative analysis between different classical algorithms, such as comparing the application effects of genetic algorithms and heuristic gradient descent methods. These limitations warrant further investigation, and future research can improve the bilevel programming model and develop various corresponding efficient algorithms.

Data Availability

All data included in this study are available upon request by contact with the corresponding author.

Disclosure

Funding agency had no involvement in the study design, data collection, analysis, interpretation of results, or writing of the manuscript.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (52232011) and the Fundamental Research Funds for the Central Universities (2682023KJ009).

References

[1] A. Ceder and N. H. Wilson, "Bus network design," *Transportation Research Part B: Methodological*, vol. 20, no. 4, pp. 331–344, 1986.

[2] O. J. Ibarra-Rojas, F. Delgado, R. Giesen, and J. Muñoz, "Planning, operation, and control of bus transport systems: a literature review," *Transportation Research Part B: Methodological*, vol. 77, pp. 38–75, 2015.

[3] S. Schéele, "A supply model for public transit services," *Transportation Research Part B: Methodological*, vol. 14, no. 1-2, pp. 133–146, 1980.

[4] P. G. Furth and N. H. Wilson, "Setting frequencies on bus routes: theory and practice," *Transportation Research Record*, vol. 818, pp. 1–7, 1981.

[5] A. F. Han and N. H. Wilson, "The allocation of buses in heavily utilized networks with overlapping routes," *Transportation Research Part B: Methodological*, vol. 16, no. 3, pp. 221–232, 1982.

[6] A. Ceder, "Bus frequency determination using passenger count data," *Transportation Research Part A: General*, vol. 18, no. 5-6, pp. 439–453, 1984.

[7] B. Capali and H. Ceylan, "A multi-objective meta-heuristic approach for the transit network design and frequency setting problem," *Transportation Planning and Technology*, vol. 43, no. 8, pp. 851–867, 2020.

[8] Z. Ahern, A. Paz, and P. Corry, "Approximate multi-objective optimization for integrated bus route design and service frequency setting," *Transportation Research Part B: Methodological*, vol. 155, pp. 1–25, 2022.

[9] J. DuráN-Micco and P. Vansteenwegen, "A survey on the transit network design and frequency setting problem," *Public Transport*, vol. 14, no. 1, pp. 155–190, 2022.

[10] S. Yoo, J. B. Lee, and H. Han, "A Reinforcement Learning approach for bus network design and frequency setting optimisation," *Public Transport*, vol. 15, no. 2, pp. 503–534, 2023.

[11] I. Constantin and M. Florian, "Optimizing frequencies in a transit network: a nonlinear bi-level programming approach," *International Transactions in Operational Research*, vol. 2, no. 2, pp. 149–164, 1995.

[12] Z. Gao, H. Sun, and L. L. Shan, "A continuous equilibrium network design model and algorithm for transit systems," *Transportation Research Part B: Methodological*, vol. 38, no. 3, pp. 235–250, 2004.

[13] B. Yu, Z. Yang, and J. Yao, "Genetic algorithm for bus frequency optimization," *Journal of Transportation Engineering*, vol. 136, no. 6, pp. 576–583, 2010.

[14] H. Spiess and M. Florian, "Optimal strategies: a new assignment model for transit networks," *Transportation Research Part B: Methodological*, vol. 23, no. 2, pp. 83–102, 1989.

[15] L. Dell'Olio, A. Ibeas, and F. RuisáNCHEZ, "Optimizing bus-size and headway in transit networks," *Transportation*, vol. 39, no. 2, pp. 449–464, 2012.

[16] H. Martínez, A. Mauttone, and M. E. Urquhart, "Frequency optimization in public transportation systems: formulation and metaheuristic approach," *European Journal of Operational Research*, vol. 236, no. 1, pp. 27–36, 2014.

[17] D. Van Lierop and A. El-Geneidy, "Enjoying loyalty: the relationship between service quality, customer satisfaction, and behavioral intentions in public transit," *Research in Transportation Economics*, vol. 59, pp. 50–59, 2016.

[18] S. Mo, Z. Bao, and B. Zheng, "Bus frequency optimization: when waiting time matters in user satisfaction," in *Proceedings of the Database Systems for Advanced Applications: 25th International Conference, DASFAA 2020*, Springer, Jeju, South Korea, September 2020.

[19] J. Cao, X. Cao, C. Zhang, and X. Huang, "The gaps in satisfaction with transit services among BRT, metro, and bus riders: evidence from Guangzhou," *Journal of Transport and Land Use*, vol. 9, no. 3, pp. 97–109, 2015.

[20] Ş İmre and D. Çelebi, "Measuring comfort in public transport: a case study for İstanbul," *Transportation Research Procedia*, vol. 25, pp. 2441–2449, 2017.

[21] Z. Xu, J. Xie, X. Liu, and Y. M. Nie, "Hyperpath-based algorithms for the transit equilibrium assignment problem," *Transportation Research Part E: Logistics and Transportation Review*, vol. 143, Article ID 102102, 2020.

[22] J. H. Wu, M. Florian, and P. Marcotte, "Transit equilibrium assignment: a model and solution algorithms," *Transportation Science*, vol. 28, no. 3, pp. 193–203, 1994.

[23] J. De Cea and E. Fernández, "Transit assignment for congested public transport systems: an equilibrium model," *Transportation Science*, vol. 27, no. 2, pp. 133–147, 1993.

[24] G. Gentile, S. Nguyen, and S. Pallottino, "Route choice on transit networks with online information at stops," *Transportation Science*, vol. 39, no. 3, pp. 289–297, 2005.

[25] C. Chriqui and P. Robillard, "Common bus lines," *Transportation Science*, vol. 9, no. 2, pp. 115–121, 1975.

[26] Q. Li, P. Will Chen, and Y. Marco Nie, "Finding optimal hyperpaths in large transit networks with realistic headway distributions," *European Journal of Operational Research*, vol. 240, no. 1, pp. 98–108, 2015.

[27] S. Nguyen and S. Pallottino, "Equilibrium traffic assignment for large scale transit networks," *European Journal of Operational Research*, vol. 37, no. 2, pp. 176–186, 1988.

[28] L. Vicente, G. Savard, and J. JúDICE, "Descent approaches for quadratic bilevel programming," *Journal of Optimization Theory and Applications*, vol. 81, no. 2, pp. 379–399, 1994.

[29] L. Armijo, "Minimization of functions having Lipschitz continuous first partial derivatives," *Pacific Journal of Mathematics*, vol. 16, no. 1, pp. 1–3, 1966.

[30] R. L. Tobin and T. L. Friesz, "Sensitivity analysis for equilibrium network flow," *Transportation Science*, vol. 22, no. 4, pp. 242–250, 1988.

[31] H. Yang and M. G. Bell, "Sensitivity analysis of network traffic equilibrium revisited: the corrected approach," in *Proceedings of the 4th IMA International Conference on Mathematics in TransportInstitute of Mathematics and its Applications*, Oxford, UK, January 2007.

[32] M. Du and A. Chen, "Sensitivity analysis for transit equilibrium assignment and applications to uncertainty analysis," *Transportation Research Part B: Methodological*, vol. 157, pp. 175–202, 2022.

[33] Y. Nie, *A Programmer's Manual for Toolkit of Network Modeling (Tnm)*, University of California, Davis, CA, USA, 2006.

[34] W. Szeto and Y. Jiang, "Transit assignment: approach-based formulation, extragradient method, and paradox," *Transportation Research Part B: Methodological*, vol. 62, pp. 51–76, 2014.

[35] W. Huanghn, "A method for bus operation cost calculation based on multi-source data," *Journal of Transport Information and Safety*, vol. 31, no. 6, Article ID 6r10, 2013.