OPTIMAL DESIGN OF SKIP-STOP TRANSIT SERVICE IN A CORRIDOR UNDER HETEROGENEOUS DEMAND

YU MEI

PhD

The Hong Kong Polytechnic University

2019

The Hong Kong Polytechnic University

Department of Electrical Engineering

OPTIMAL DESIGN OF SKIP-STOP TRANSIT SERVICE IN A CORRIDOR UNDER HETEROGENEOUS DEMAND

YU MEI

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

July 2018

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____(Signed)

Mei Yu _____(Name of student)

Abstract

Skip-stop service, also termed "limited-stop service", is a transit service with more than one transit route operating simultaneously on the same line and each route only visiting a subset of the stops. Compared to the conventional all-stop service (where a transit vehicle visits each and every stop along the line), the skip-stop service has higher commercial speed and patrons can enjoy reduced in-vehicle travel time. This service scheme has long been both studied in the literature and implemented in real cities. The majority of the previous studies have formulated discrete models for optimizing the skip-stop routing plan only; i.e., they optimize the selection of stops to be visited by each skip-stop route from a given set of stops. Hence, they fail to jointly optimize the skip-stop routing plan and stop locations. In addition, those discrete models were often solved by heuristic methods. Thus evaluating the solution quality, in terms of the optimality gap between the heuristic solution and the global optimum, is difficult. To address these deficiencies, this thesis has developed continuum approximation (CA) models to optimize various forms of skip-stop services designs under spatially heterogeneous demand. Efficient solution algorithms are also proposed and tested via extensive numerical examples.

Specifically, the following three forms of skip-stop services are considered: i) AB-type service; ii) local-express service; and iii) a general form of skip-stop service.

In an AB-type service, different skip-stop routes visit the non-transfer stops in a rotating fashion. If there are only two routes named route A and route B, then the non-transfer stops will be placed in an "ABABAB…" fashion; hence the term AB-type service. Each transfer stop will be visited by all the skip-stop routes. We have formulated CA models for jointly optimizing the stop spacings, the number of skip-stop routes, and the transfer stop spacings under any heterogeneous demand patterns. Near-optimal solutions to the CA formulation are obtained via an efficient solution approach, employing the calculus of variations method in an iterative algorithm. A discretization recipe is also proposed to convert the CA solution to a real design plan. Numerical case studies confirm the practicality of the models, the efficiency of the solution approach, and the advantages of the AB-

type design over the conventional all-stop design under various heterogeneous demand patterns.

A local-express service consists of an express route and a local route. The express route only visits the express stops (which are also transfer stops between the express and local routes), and the local route visits all the stops. CA models are again formulated for optimally designing the express and local stop spacings under heterogeneous demand. The models explicitly account for passengers' choices between different route options (for example, taking the local route only or taking the local route as feeder to access the express route). Numerical case studies compared the optimal local-express design with other corridor designs.

Finally, a more general skip-stop service is also studied in this thesis, where the non-transfer stops of different routes and the transfer stops can both be distributed along the line in an arbitrary fashion. A novel formulation for optimizing this general design has been developed in this thesis, which is discrete in nature but inspired by the conventional CA models. Specifically, our model allows the stops to be located anywhere along the line, with stop densities instead of individual stop positions employed as decision variables. Efficient solution methods for this new formulation yielding near-optimal solutions under various heterogeneous demand patterns are presented. Comparison between different skip-stop design forms shows that the general skip-stop design is a generalization of the AB-type design and under some demand patterns the general design can significantly outperform AB-type and local-express designs.

Acknowledgements

First and foremost, I would like to express my sincere appreciation and gratitude to my chief supervisor, Dr. Weihua Gu, for his guidance, support and care in both research and daily life. He is a brilliant, vigorous and responsible mentor, who has done everything to help me learn how to develop research ideas, write convincing papers, and improve presentations. He is also a patient and trusted friend, always giving real help whenever I was in trouble. The knowledge, skills, and ways of thinking that I have learned from him have become my precious treasure. It is my great honor and fortune to be his student.

I shall deliver the same enormous thanks to Dr. Wenbo Fan, a close collaborator of my thesis research. He is a great advisor and a good friend. I am highly impressed by his integrity, passion and creativity. The days and nights working with him are unforgettable.

My acknowledgements also go to other PolyU faculty members for their guidance and support throughout my studies here. I would like to express my sincere gratitude to Prof. Alan Lau, Prof. Edward Chung, and Dr. Jinwoo Lee, your inspiring ideas and insightful comments have greatly improved my research.

My appreciation is also given to my good friends and fellows at PolyU: first to Joy, who is like my sister and always encouraged me when I got lost in research work. Together we studied how to develop mathematic formulations, learned new solution techniques and presentation skills, …, too many to list here. It is always a great joy to work with you. I also want to thank Samuel, Christine, Xiao Yang, Fuliang Li, Dr. Xin Li, Fangyi Yang, Dr. Sangen Hu, Yuan Lu and many other colleagues and friends, for their sincere care and friendship.

Lastly, I extend special thanks to my parents Lixin Mei and Lihua Qian, and my wife Fan Gu. Words are too short to express my gratitude for your selfless love and faithful support. Having one of you is already a great fortune, and I have you all. You are always my love, forever…

Table of Contents

Chapter 1 Introduction

Section 1.1 presents the background of this research and Section 1.2 reviews related studies in the literature. Section 1.3 provides a dissertation overview.

1.1 Background

Public transport corridors play a vital role in urban areas for efficiently serving cross-district mass travel demand (e.g., commuters). These corridors are crucial for densely populated metropolitan areas such as New York City, San Francisco, London, Tokyo, Beijing, and Hong Kong. For instance, The Metro Line 1 in Beijing and Metro Line 2 in Shanghai each provide more than 1 million passenger trips daily. Hence, improving the transit service in these essential corridors is of great importance for both city managers and transit patrons. They are the cardio vascular system supporting the life blood and breathe of the world's greatest cities.

To better serve vast passenger numbers, skip-stop service schemes are often employed in heavy-traffic transit corridors. Skip-stop service, also termed "limited-stop service", is defined as a transit service where more than one transit routes operate simultaneously on the same line, with each route visiting only a subset of the stops. This service scheme has long been implemented in real cities and studied in the literature. Compared to the conventional all-stop service (where a transit vehicle visits each and every stop along the line), the skip-stop service can operate at a higher commercial speed, save patrons' travel time, improve the productivity of transit fleets, and increase transit schedule reliability.

Skip-stop service operates in a number of specific forms, including: i) zonal service, where a transit vehicle starts the service run by visiting all the stops within a zone, and then performs a non-stop line-haul travel to the city center; ii) short-turn service, where some service runs only cover a continuous portion of the line and skip the remaining part of the line entirely; iii) local-express service, where an all-stop local line operates in parallel with an express line which only stops at a few express stops; iv) AB-type service (Vuchic, 2007), where two or more lines operate in parallel in a corridor, each visiting all the transfer stops and a subset of the non-transfer stops (which are arranged in a rotating fashion; see the detailed illustration in Section 1.3); and v) skip-stop services of the general form,

where multiple skip-stop lines simultaneously serve a corridor, each visiting an arbitrary subset of stops.

Zonal and short-run services are perhaps most commonly used in the real world due to their simplicity (Larrain et al. 2015; Cortes et al. 2011). For example, the Massachusetts Bay Transit Authority (MBTA) has been operating a short-run service (bus route 57A) in the Watertown-Brighton corridor in the Boston metropolitan area during weekday rush hours since 2008. This short-run service effectively mitigates overcrowding along the middle part of the route (Belcher, 2015). Zonal services have also been operating in the same corridor to connect the Downtown Boston and suburban areas. These services are relatively easy, both for the agency to design and implement, and for the patrons to understand and use. However, their applicability is often limited to certain specific demand patterns; e.g., zonal service is suitable for the many-to-one demand pattern, and short-run service is suitable for corridors where demand is high in the middle section and low in the other parts.

More complicated forms of skip-stop service are also operated in various cities. For example, the local-express service has been used in the San Pablo Avenue of San Francisco Bay Area (Mejias and Deakin, 2005), the TransMilenio system in Bogota, Colombia (Hidalgo et al., 2013) and New York City (Hirsch et al., 2000). The TransSantiago Metro in Santiago, Chile, has implemented AB-type service on line 2, line 4, and line 5 during daily rush hours since 2007 (Metro de Santiago, 2008). This AB-type service scheme can be traced back to 1947, when the Chicago Metro system introduced possibly the world's first AB-type service (with two routes) to some rail corridors after the design failure of a very complicated local-express service (Chicago-L, 2010). A great advantage of the AB-type service is that it is easy for the patrons to understand and use and it does not need a second track for trains to bypass each other. Other forms of skip-stop services (such as the local-express service) often require a resource expensive second track or passing lane to operate.

Other, more general, skip-stop service forms have also been studied in the literature (see Section 1.2.1). However, these have few real-world applications in urban transit systems, partly because these schemes are more difficult to design and operate, and can be confusing for patrons to use.

1.2 Literature Review

Section 1.2.1 describes the studies of different schemes of skip-stop service in urban transit systems. Section 1.2.2 discusses two solution method categories related to optimal skip-stop service design.

1.2.1 Skip-stop service schemes in urban transit systems

Turnquist (1979), presented an early study of zonal service aimed at minimizing patrons' travel time which employed a dynamic programming method to optimize the number of zones, the boundaries of zones and the allocation of buses within a fixed fleet size. This work was extended by Jordan and Turnquist (1979) who also considered bus travel time reliability. Their model can minimize either average travel time or travel time variance. Later, Ghoneim and Wirasinghe (1986) investigated the benefits of operating zonal service for an existing rail line. Their model incorporated some constraints to ensure minimal headway between trains on a single track and optimized a more general objective function consisting of both the patrons' cost and the operating cost. Furth (1986) presented a method to combine the zonal service design with all-stop service and extended the use of zonal service from linear corridors to a branching corridor.

The above works sought the optimal design of zonal service for existing rail systems or bus systems in which stop locations are considered as given. Ghoneim and Wirasinghe (1987) included the stop spacings as decision variables and addressed the problem of designing zonal service for newly planned urban commuter rail lines. Recently, Larrain et al. (2015) have generalized the definition of zonal service by specifying that the terminal is not necessarily a single stop but can also be a zone. They presented two heuristics to optimize the design of this zonal service in congested and uncongested cases.

Overall, these works have shown that the zonal service can outperform the all-stop service significantly under the many-to-one demand pattern.

Short-turn service is usually operated together with regular all-stop service. It is particularly useful in corridors where the demand is high in a continuous portion of the corridor and low in other parts of the corridor. Furth (1987) pointed out that schedule coordination between short-turn vehicles and all-stop vehicles is

essential for balancing passenger loads and minimizing costs. He presented an algorithm to find the optimal headways and locations of turn-back points in order to minimize fleet size. Delle Site and Filippi (1998) addressed the stochasticity in bus arrivals and developed a multi-period short-turn service, which considered both elastic and inelastic demands. This problem was formulated as a maximization of social benefits for the case of elastic demand, and a minimization of total costs for the case of inelastic demand. Later, Tirachini et al., (2011) presented a short-turn design model to optimize user and operator costs and derived analytical expressions for optimal headways and turn-back points. These analytical expressions were used to analyze the influence of different parameters on short-turn service performance. They found that the benefit of short-turn service hinged on how the demand was spatially concentrated: the more spatially concentrated the trips are, the greater the benefit would be. Cortes et al. (2011) further explored integrating short-turn service with dead-heading (dead-heading means empty vehicles return to the line's start point in the low-demand direction to increase service frequency in the other, high-demand direction). They found that the short turning strategy generally yielded greater benefit than dead-heading due to the extra costs of operating empty vehicles.

In zonal and short-turn services, the skip-stop vehicles only serve the patrons that arrive at a continuous portion of the corridor. On the other hand, in more general local-express services, an express line serves selected stops throughout the corridor in parallel with an all-stop local line. A key issue when designing a local-express service is, therefore, to model the assignment of patrons between the local and express lines: a patron many choose to take either the local line only to complete her trip, or the local line first and later transfer to an express line. Incorporating patrons' route choice imposes greater modeling challenges. Hence, some studies assumed that no transfer is allowed in an attempt to simplify the modeling work (e.g. Chen, X. et al., 2015; Chiraphadhanakul and Barnhart, 2013; Liu et al., 2013, Thilakaratne and Wirasinghe, 2016).

Ulusoy et al. (2010) is one of the few studies that explicitly modeled patron route choices by developing a logit model to assign the demand to different route options. An iterative process was employed to derive the equilibrium of demand

assignment. Their model also integrated the short-turn service into the local-express design.

The AB-type service form is also limited because the non-transfer stops are arranged in a rotating fashion. However, this service form also has advantages. First, different routes' travel times over the same portion of the corridor are similar, and thus patron route choice behavior and patron decision making, can be simplified. Secondly multiple routes can operate along the same corridor without incurring overtaking maneuvers between transit vehicles, even when the headways between consecutive vehicles are small. This is especially useful for single-track urban rail systems serving a high demand.

Freyss et al. (2013) studied how to optimize the headway and routing plan of this service for a two-route case. They assumed that demand was uniformly distributed along the corridor and explored the relationship between cost and stop spacing. Lee et al. (2014) presented a model to optimize the AB-type service under heterogeneous demand patterns and proposed a genetic algorithm to solve the problem.

Abdelhafiez et al. (2017) summarized the variants of the AB-type service, and classified them into five strategies according to the way to arrange the non-transfer stops, namely pairing, alternating, modified alternating, non-backtracking, and free assignment. They proposed different models for these strategies to minimize the patrons' travel time, and developed a heuristic algorithm for solving these models. They found that the average patron's travel times under these different strategies are quite similar.

System design involving more than two routes has been investigated by Gu et al. (2016). That work also incorporated optimization of stop locations and coordination between different skip-stop routes. However, similar to Freyss et al. (2013), Gu et al. also assumed that demand was uniformly distributed. Thus the applicability of their models in real practice is quite limited.

All the above service forms are special cases of the general skip-stop service design, which allows multiple skip-stop routes to simultaneously serve a corridor, each visiting an arbitrary subset of stops. Optimizing this general form of skip-stop service is much more difficult than optimizing the special forms due

to the much larger solution space of the routing plan. For this reason, the optimal design problem of more general skip-stop services was often formulated as a two-stage problem in the literature: one stage optimizing the service headways and the other optimizing the skip-stop routing plan. Patron route choice behaviors were usually simplified, for example by assuming that patrons will always choose to take the first vehicle that arrives at the stop. For instance, Levia et al. (2010) presented a model to optimize only the headways of the skip-stop service with vehicle capacity constraints under a given skip-stop routing plan. Larrain et al. (2013) proposed a few heuristic methods to generate the skip-stop routing plan. Chen, J. et al. (2015) further extended the work of Liu et al. (2013) to a multiple route design and proposed a hybrid artificial bee colony and Monte Carlo algorithm to find the optimal skip-stop routing plan. Niu et al. (2011) studied the optimization of both headways and routing plan under time-dependent demands and proposed a bi-level genetic algorithm to iteratively determine the headways and the routing plan.

Soto et al. (2017) is among the few that have embedded route assignment models into the optimization problem. In their modeling framework, headway optimization and route assignment are iteratively processed. Two route assignment models, i.e. a deterministic model and a stochastic model, were examined. The results showed that the stochastic model led to more realistic and robust solutions, although this made the problem much harder to solve.

1.2.2 Solution methods

Most works in the realm of optimal skip-stop service design for urban transit systems have relied on discrete models. They take the discrete OD demand as inputs and use binary variables to represent if a stop is skipped by a route or not. These works took headways and skip-stop routing plan as the decision variables, while assuming that stop locations along the line are given. Their objective is either to minimize the patrons' total travel time or to minimize the generalized cost of the system, which is the sum of the patrons' travel time cost and the agency's operating cost.

Exact solution methods have been developed for simple forms of skip-stop services. For instance, Turnquist (1979) and Jordan and Turnquist (1979)

employed the dynamic programming method to optimize zonal service. Tirachini et al., (2011) and Cortes et al. (2011) derived analytical solutions for the optimization of short-turn design that minimizes the total user and operator costs. Exact methods have also been applied to skip-stop service optimization in intercity rail systems (Yang et al., 2016; Qi et al. 2018; Yue et al, 2016; Jiang et al. 2017). Note that intercity rail systems' optimization objective is usually different from that of urban transit systems. Specifically, the nonlinear cost terms that are necessary for urban transit systems were ignored in intercity rail models, including the patrons' access/exit cost, in-vehicle travel cost, and certain agency cost terms. Due to that, the models for skip-stop service in intercity rail systems can often be formulated as mixed-integer programs, and exact methods can be developed.

However, for the more complicated skip-stop service designs (e.g. local-express, AB-type, and more general forms) in urban transit systems, exact methods are usually unavailable. This is partly due to the nonlinear and non-convex features of the formulation, and the problem's large solution space. The optimal design of the skip-stop routing plan on a medium-sized corridor would involve dozens or even hundreds of binary decision variables, not to mention that even more variables will be included if the stop locations are jointly optimized. Hence, studies of this kind have to rely on heuristic methods, such as genetic algorithm (Liu et al. 2013; Lee et al. 2014; Niu et al. 2011; Chen, X. et al. 2015), tabu search (Li et al., 1995), artificial bee colony (Chen, J. et al. 2015), or iterative methods (Ulusoy et al. 2010; Niu et al. 2011; Soto et al. 2017). Therefore, assessing the quality of those solutions (how close they are to the global optima) is often difficult.

Note furthermore that most works in the literature, including all those cited above, have assumed that the candidate stop locations were given so only the routing plan and the vehicle dispatch schedules were optimized. To our best knowledge, no discrete model has been proposed for jointly optimizing stop locations along a corridor in combination with the skip-stop routing plan.

In contrast to the discrete models, continuous approximation (CA) models are based upon a number of continuous functions, which approximate the temporally- or spatially-varying discrete parameters and variables (for example the OD demand, line and stop spacings) of a transit system. The CA approach has been used for developing models for optimizing transit system designs, including

the line and stop spacings (Holryod, 1967; Newell, 1971, 1973, 1979; Wirasinghe and Ghoneim, 1981; Vaughan, 1986; Daganzo, 2010; Chen et al., 2015; Gu et al., 2016).

Vaughan and Cousins (1977) appears to be the first work on developing CA models for a transit route, where the demand is assumed to be continuous and follows a many-to-many demand pattern. Stop density (the reciprocal of stop spacing) is chosen as the decision function, which is location-dependent along the route. The optimization problem (the minimization of transit users' travel time) is solved using calculus of variations. Later, Wirasinghe and Ghoneim (1981) proposed a more general CA-based model for determining bus stop spacing (also expressed as a function of location) along a transit route section. Li and Bertini (2009) simply assumed a uniform stop spacing and optimized that for a bus route using a 1-year demand dataset collected in Portland, Oregon (US). Recently, Medina et al. (2013) have applied a similar CA-based transit design model to a bus route considering time-varying demand in Santiago, Chile. In particular, they considered more realistic conditions in their optimal design such as asymmetric stop locations between the two travel directions, the passenger-carrying capacity of vehicles, and the vehicle-carrying capacity at the stops.

Some works extended the design of a single transit route to more complicated line structures along a corridor. For example, Wirasinghe et al. (1977), Hurdle and Wirasinghe (1980), and Wirasinghe (1980) studied the optimal design of a rail corridor with feeder transit service (e.g., feeder buses). In Wirasinghe et al.'s work (1977), the rail-bus corridor (trunk-feeder corridor in general) was optimized to maximize the users' benefit, and the results compared against a scenario of providing direct-bus service only along the corridor. Their decision variables include the rail station spacing (as a function of location) and the train headway during peak period. Hurdle and Wirasinghe (1980) examined the impacts of several feeder transit modes (walking, bus, park-and-ride, and bicycle) on the optimal design solutions of the trunk-feeder corridor. Further, Wirasinghe (1980) jointly optimized the route density (also expressed as a function of location) and headways of the feeder buses along a rail corridor.

Thanks to the nice mathematical properties of most CA models in the literature, analytical formulas have been developed as the optimal solution, or

large part of the solution, in all of the above-cited CA works. By virtue of these analytical results, the CA-based transit design models enjoy computational efficiency, allowing researchers to conduct extensive parametric analysis to unveil more findings on the cause-and-effect relationship between key input parameters and optimal design outcomes. However, these optimal solutions, often cannot be directly applied in practice due to their idealized and unrealistic assumptions (such as assuming continuous demand) and the continuum approximation made in the formulations. Fortunately, these solutions are often robust to small changes in the operating parameters, and the optimization objective is usually insensitive to small adjustments of the solution in the neighborhood of the optimum (for example adjusting individual stop placement locally to avoid junctions or ramps). This is because the CA objective functions have similar forms as the Economic-Order-Quantity (EOQ) model, which is known to have robust solutions (Estrada et al., 2011). As such, the optimal CA solutions can provide transit planners a rough, high-level picture about how a transit system should be laid out. Subsequently, the solution can be fine-tuned to obtain realistic design plans.

In summary, most previous skip-stop service design studies formulated discrete models for optimizing the service headway and routing plan. These discrete models were often solved by heuristic methods but evaluating the solution quality in terms of the optimality gap between the heuristic solution and the global optimum, is often difficult. In those discrete model works, optimal designs that also involve stop location optimization were only studied for zonal services due to their simple structure. For more complicated skip-stop service structures, the CA approach appears to be a promising way to jointly optimize stop locations and the skip-stop routing plan. Freyss et al. (2013) and Gu et al. (2016) pioneered this research direction by examining the optimal designs for AB-type and local-express services but only under uniformly-distributed demand. However, to our best knowledge no examination of this joint design problem under more realistic heterogeneous demand patterns has been conducted until now.

1.3 Dissertation Overview

This dissertation presents the study of the following three forms of skip-stop services under spatially heterogeneous demand patterns: i) AB-type service; ii)

local-express service; and iii) a more general form of skip-stop service. CA models are developed for optimizing the joint design of these services, including their stop locations, routing plans, and service headways. Efficient solution algorithms are also developed and tested via extensive numerical examples.

The dissertation is organized as follows:

Chapter 2 examines the AB-type service. A 3-route example of this service scheme is illustrated in Figure 1.1, where the three routes are labeled A, B, and C. The stops are classified as non-transfer stops (marked by squares) or transfer stops (marked by black dots). Non-transfer stops labeled A, B, and C will only be visited by route A, B, and C, respectively. Transfer stops will be visited by all three routes. The line segment between two consecutive transfer stops is termed a 'skip-stop bay'. Within each skip-stop bay, stops of route A, B, and C appear in turn for k times ($k = 2$ for the example in Figure 1a).Our model follows Freyss et al. (2013) and Gu et al.'s (2016) logical assumption that a patron will always access and egress the transit system through the nearest stops to her origin and destination. Hence, a patron has to make a transfer if her origin and destination stops belong to different routes. Note that if a trip's origin and destination stops belong to different routes, but are located in the same skip-stop bay, then the trip will contain a backtracking segment, as illustrated in Figure 1.1.

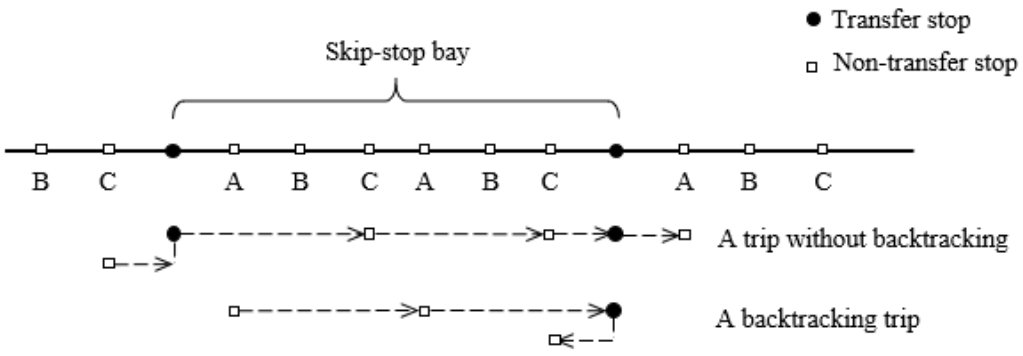


Figure 1.1 The AB-type service design

CA models of the AB-type service under heterogeneous demand patterns are formulated in Chapter 2, aiming at minimizing the total generalized costs for both patrons and the operating agency. A near-optimal solution approach is proposed using the calculus of variations method. This solution is then validated

by examining the gap between the resulting generalized cost and the optimal generalized cost of a proposed lower bound problem.

A discretization recipe is also proposed to convert the CA solution to a real design plan where exact stop locations are given. The advantages of this AB-type design over the conventional all-stop design has been illustrated by a large number of numerical cases.

Chapter 3 examines local-express service. This service design is illustrated in Figure 1.2, where one local line and one express line are planned. Local stops are marked by squares, and express stops are marked by black dots. The local line will visit all the stops and the express line will only visit express stops. Although patrons may have multiple route choices, they are assumed to always choose the route with the shortest travel time. For instance, a patron travelling from a local stop to another a local stop can choose either to take a local line only (without any transfers) or to take a local line first, then transfer to an express line, and finally transfer back to the local line to reach her destination, whichever has shorter travel time. The two kinds of trips are illustrated in Figure 1.2.

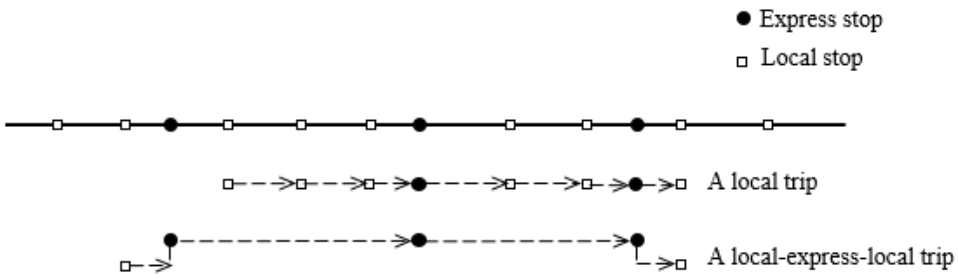


Figure 1.2 The local-express service design

CA models of the local-express service are formulated again to minimize the generalized cost of the system. The decision variables include the headways and stop spacings for both local and express lines. Our model allows for local and express lines using differentiated transit technologies (such as rail for the express line and bus for the local line). A three-stage iterative algorithm to solve the model and a discretization recipe to convert the CA solution to a real design plan is proposed. Numerical analysis also shows that the optimal local-express design outperforms the conventional all-stop design under various operating conditions.

Finally, in Chapter 4 a more general skip-stop service is examined. As illustrated in Figure 1.3, now the non-transfer stops of different routes can be distributed along the line in an arbitrary fashion. This design is quite general since AB-type service and local-express service can both be considered as special cases within it. However, for simplicity, some AB-type service assumptions are taken in this chapter. These assumptions include: i) that each skip-stop route visits all the transfer stops; ii) a patron always accesses the transit system through the nearest stop to her origin, and egresses through the nearest stop to her destination; and iii) at most one transfer was involved in any patron's trip. Hence, patron route choices are simplified in our modeling work.

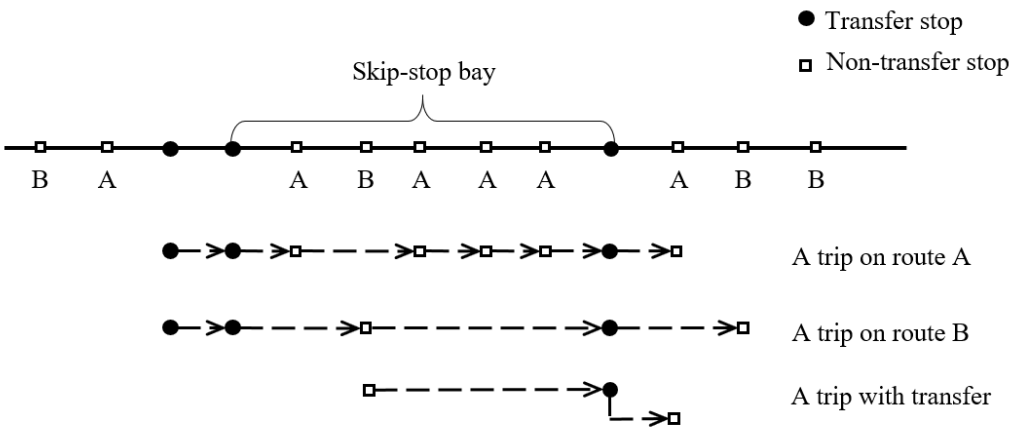


Figure 1.3 A more general skip-stop service design

Our novel formulation to model this highly flexible general design combines the properties of discrete models and CA models. Our model takes a discrete OD matrix as input, and uses stop densities instead of individual stop locations as decision variables. A heuristic method is proposed to solve the problem. A relaxed problem, which is much easier to solve than the original problem, is developed to: i) provide a good initial solution to enhance the heuristic method; and ii) provide a lower bound which can be used to examine the heuristic solution's quality. A discretization recipe is also provided. Numerical analysis verifies that this model can generate near-optimal designs under various demand patterns. Extensive numerical results reveal that the general skip-stop design outperforms the local-express design, and it also outperforms the AB-type design if the extra infrastructure cost for a second track or bus lane is not considered. Note

that the AB-type design does not require a passing track or bus lane, while both local-express and general form designs do.

Chapter 5 concludes this dissertation and discusses potential extensions of the present research.

Chapter 2 AB-type Designs

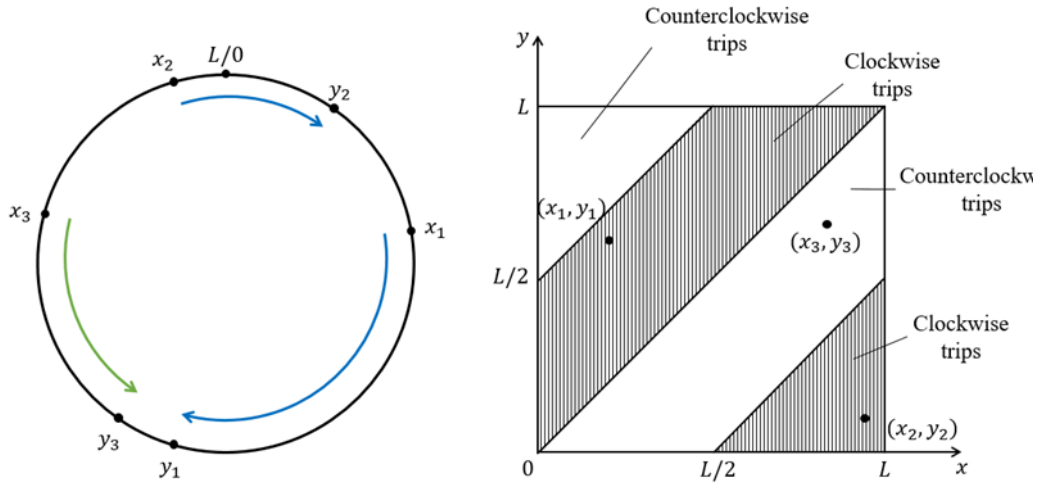
This chapter presents CA models developed to optimize the AB-type service in a loop corridor. The models jointly optimize the stop locations and the routing plan under arbitrary heterogeneous demand pattern. Unlike most previous CA models in the literature, the optimization formulation for the heterogeneous AB-type design cannot be fully decomposed by the local spatial coordinate. Nevertheless, we show that this model can still be efficiently solved to near-optimality by integrating the calculus of variations method into an iterative process. How the CA solution can be converted to a real corridor design with discrete stop locations and routing plan is also described. The solution algorithms' optimality gap is examined via a lower bound through a large array of numerical examples. Numerical analysis also shows that the optimal AB-type service design outperforms the conventional all-stop design under various operating conditions.

Section 2.1 presents the formulations of the AB-type design. Section 2.2 describes the proposed models' solution method. Section 2.3 provides numerical analysis. Section 2.4 summarizes the AB-type design models.

2.1 The CA Formulations

We consider a bi-directional transit service operating along a loop corridor of length L as shown in Figure 2.1a. (A linear corridor can be modeled in a similar fashion, but the modeling details would be different.) The demand is assumed to be exogenous, and its density (in the unit of trips/km^2/h) is represented by a slow-varying, integrable function $\lambda(x, y)$, where x and y are location coordinates of the trip origin and destination, respectively. The x and y are measured along the corridor in the clockwise direction; $0 < x, y \leq L$. We assume that a patron always chooses the direction of service that minimizes her travel distance; i.e. the patron takes the clockwise service if $0 < y - x \leq \frac{L}{2}$ or $y - x \leq -\frac{L}{2}$, and the counterclockwise service if $-\frac{L}{2} < y - x \leq 0$ or $y - x > \frac{L}{2}$. Figure 2.1b shows a square in which each point represents a possible OD pair in the corridor. The ranges of OD pairs for clockwise trips are marked by shading. Two clockwise trips

((x_1, y_1) and (x_2, y_2)) and one counterclockwise trip ((x_3, y_3)) are also illustrated in Figure 2.1a and b.



(a) The loop corridor and three example trips.

(b) Ranges of clockwise and counterclockwise OD pairs.

Figure 2.1 Clockwise and counterclockwise trips in a loop corridor.

For convenient development of our models, some aggregate demand functions, such as the density function of trip origins, need to be defined first. We first define the *left δ-neighborhood* of x in our loop corridor as $U_-^L(x, \delta) \equiv [\max(x - \delta, 0), x] \cup [\min(x - \delta + L, L), L]$. The $U_-^L(x, \delta)$ is a set of all the points that can be reached by traveling from x counterclockwise for a distance no more than δ. Specifically, $U_-^L\left(x, \frac{L}{2}\right)$ represents the set of destinations for counterclockwise trips originating from x, and $U_-^L\left(y, \frac{L}{2}\right)$ represents the set of origins for clockwise trips destined for y. Similarly, we define the *right δ-neighborhood* of x as $U_+^L(x, \delta) \equiv [x, \min(x + \delta, L)] \cup [0, \max(0, x + \delta - L)]$, which indicates a set of all the points that can be reached by traveling from x clockwise for a distance of no more than δ. The $U_+^L\left(x, \frac{L}{2}\right)$ represents the set of destinations for clockwise trips originating from x, and $U_+^L\left(y, \frac{L}{2}\right)$ represents the

set of origins for counterclockwise trips destined for y.[1] We further define the δ-*neighborhood* of x as $U^L(x,\delta) \equiv U_-^L(x,\delta) \cup U_+^L(x,\delta)$.

Built upon the above definitions, the following aggregate demand functions and variables are defined:

i) The densities of origins for clockwise and counterclockwise trips at location x, denoted by $P_c(x)$ and $P_{cc}(x)$, respectively; and the densities of destinations for the two types of trips at location y, denoted by $Q_c(y)$ and $Q_{cc}(y)$ respectively:

$$P_c(x) = \int_{y \in U_+^L\left(x,\frac{L}{2}\right)} \lambda(x,y)dy\,,\ 0 < x \le L \tag{2.1a}$$

$$P_{cc}(x) = \int_{y \in U_-^L\left(x,\frac{L}{2}\right)} \lambda(x,y)dy\,,\ 0 < x \le L \tag{2.1b}$$

$$Q_c(y) = \int_{x \in U_-^L\left(y,\frac{L}{2}\right)} \lambda(x,y)dx\,,\ 0 < y \le L \tag{2.1c}$$

$$Q_{cc}(y) = \int_{x \in U_+^L\left(y,\frac{L}{2}\right)} \lambda(x,y)dx\,,\ 0 < y \le L \tag{2.1d}$$

ii) The total clockwise and counterclockwise demand, denoted by Λ_c and Λ_{cc}, respectively:

$$\Lambda_c = \int_{x=0}^{L} P_c(x)dx = \int_{y=0}^{L} Q_c(y)dy \tag{2.2a}$$

$$\Lambda_{cc} = \int_{x=0}^{L} P_{cc}(x)dx = \int_{y=0}^{L} Q_{cc}(y)dy \tag{2.2b}$$

iii) The (approximate) flows of on-board patrons in the clockwise and counterclockwise directions at location x, denoted by $o_c(x)$ and $o_{cc}(x)$, respectively:

$$o_c(x) = \int_{z \in U_-^L\left(x,\frac{L}{2}\right)} \int_{y \in U_+^L\left(z,\frac{L}{2}\right) \cap U_+^L\left(x,\frac{L}{2}\right)} \lambda(z,y)dzdy \tag{2.3a}$$

---

[1] For a linear corridor, the δ-neighborhoods are defined as $U_-^L(x,\delta) = [\max(x-\delta,0),x]$ and $U_+^L(x,\delta) = [x,\min(x+\delta,L)]$. The $U_-^L(x,L)$ represents the set of destinations for "leftward" trips originating from x, and $U_-^L(y,L)$ represents the set of origins for "rightward" trips destined for y; similarly, the $U_+^L(x,L)$ represents the set of destinations for "rightward" trips originating from x, and $U_+^L(y,L)$ represents the set of origins for "leftward" trips destined for y.

$$o_{cc}(x) = \int_{z \in U_{+}^{L}\left(x,\frac{L}{2}\right)} \int_{y \in U^{L}\left(z,\frac{L}{2}\right) \cap U^{L}\left(x,\frac{L}{2}\right)} \lambda(z,y) dz dy \quad (2.3b)$$

The (2.3a-b) are approximations because the exact on-board patron flows are affected by the locations of transit stops (note that the patrons' boarding and alighting behavior only occurs at the stops).

We next present the CA models for optimizing the conventional all-stop service (Section 2.1.1) and the proposed AB-type service (Section 2.1.2). In the interest of simplicity, we adopt the following assumptions that have been commonly used in the literature (Daganzo, 2010; H. Chen et al., 2015; Gu, et al., 2016): i) a patron always accesses and egresses the transit system through the nearest stops from her origin and destination, respectively; ii) a transit vehicle spends a constant time, τ, at each stop loading and unloading patrons, including the time lost due to the vehicle's acceleration and deceleration[2]; iii) the clockwise and counterclockwise services share the same set of stops; and iv) patrons arrive randomly at their origin stops, regardless of the service schedule, which is true when the transit service frequency is high and therefore patrons often do not bother to observe the schedule.

A complete list of notation is furnished in Appendix A.

2.1.1 Formulation for the all-stop service

The models presented in this section are similar to those proposed by Wirasinghe and Ghoneim (1981). Our transit cost structure also follows the conventions in the literature (Daganzo, 2010; Sivakumaran et al., 2014; Gu et al., 2016). To provide a complete picture these are described in full detail here. The description of these simple models also forms a basis for the presentation of more complicated skip-stop models later in this dissertation.

To minimize the transit system's generalized cost, consisting of the patrons' travel cost and the transit agency's cost. The patrons' total trip cost per operation

[2] In some papers (e.g. Cipriani et al., 2012; Meng and Qu, 2013), a bus's dwell time at a stop was assumed as a linear function of the number of boarding patrons at the stop. With modest changes, the methodology presented in this paper can still be applied if this alternative assumption is used instead.

hour is the sum of the following components: i) the total access and egress time by walking, AUC_a; ii) the total wait time, AUC_w; and iii) the total in-vehicle travel time, AUC_i. These cost terms are formulated as follows:

$$AUC_a = \int_{x=0}^{L} \frac{s(x)}{4v_w}\left(P_c(x) + Q_c(x) + P_{cc}(x) + Q_{cc}(x)\right)dx \tag{2.4}$$

$$AUC_w = \frac{H_c}{2}\Lambda_c + \frac{H_{cc}}{2}\Lambda_{cc} \tag{2.5}$$

$$AUC_i = \int_{x=0}^{L}\left(o_c(x) + o_{cc}(x)\right)\left(\frac{1}{v} + \frac{\tau}{s(x)}\right)dx \tag{2.6}$$

where the decision variables/functions include: $s(x)$, which denotes a continuum approximation of stop spacing as a function of location x ($0 < x \leq L$); and H_c and H_{cc}, which denote the transit service headways in clockwise and counterclockwise directions, respectively. The operating parameters are: v_w which denotes the patrons' walking speed; and v, which denotes the transit vehicles' cruise speed. Equation (2.4) is derived from the fact that the average access or egress walking time of a patron originating or destined for x is $\frac{s(x)}{4v_w}$. Equation (2.5) is based upon the fact that the average wait time per patron is half of the service headway.[3] The right-hand-side of (2.6) is an integral of the total patron in-vehicle travel time accrued in a unit distance at location x, which is the product of the total on-board flow, $o_c(x) + o_{cc}(x)$, and the average vehicle travel time per unit distance, $\frac{1}{v} + \frac{\tau}{s(x)}$ (τ is dwell time at a stop).

The agency's cost per operation hour consists of: i) the distance-based vehicle operating cost (mainly fuel cost) AAC_K; ii) the time-based vehicle operating cost (including for example amortized vehicle purchase cost and staff wages), AAC_H; iii) the amortized line infrastructure cost (such as for busways or rail tracks), AAC_I; and iv) the amortized stop infrastructure cost, AAC_S. These are formulated as follows:

[3] For simplicity, here we assume the headway is a constant. Headway variations caused by transit service instability (Newell and Potts, 1964) have a second-order effect on the generalized cost and are thus omitted in this paper. Please refer to Daganzo (2009) for control methods to stabilize the headways.

$$AAC_K = \frac{\pi_v L}{\mu}\left(\frac{1}{H_c} + \frac{1}{H_{cc}}\right) \tag{2.7}$$

$$AAC_H = \left(\frac{1}{H_c} + \frac{1}{H_{cc}}\right)\frac{\pi_m}{\mu}\int_{x=0}^{L}\left(\frac{1}{v} + \frac{\tau}{s(x)}\right)dx \tag{2.8}$$

$$AAC_I = \frac{2\pi_i L}{\mu} \tag{2.9}$$

$$AAC_S = \frac{\pi_s}{\mu}\int_{x=0}^{L}\frac{1}{s(x)}dx \tag{2.10}$$

where π_v and π_m are the unit operating costs per vehicle-km and per vehicle-service hour, respectively; π_i and π_s are the unit construction and maintenance costs per km of line infrastructure (per direction) and per stop, respectively, amortized for each hour of operations; μ denotes the value of time for the patrons, which is a proxy of the average hourly wage rate among the patrons.

The generalized cost minimization problem for the all-stop service can thus be formulated as follows:

$$\min_{s(x),H_c,H_{cc}} AC = AUC_a + AUC_w + AUC_i + AAC_K + AAC_H + AAC_I + AAC_S \tag{2.11a}$$

subject to:

$$H_{min} \le H_c \le \frac{K}{\max\limits_{0<x\le L}\{o_c(x)\}} \tag{2.11b}$$

$$H_{min} \le H_{cc} \le \frac{K}{\max\limits_{0<x\le L}\{o_{cc}(x)\}} \tag{2.11c}$$

$$s(x) > 0 \tag{2.11d}$$

where H_{min} denotes the minimum headway due to technical or safety constraints (Gu et al., 2016); and K denotes a transit vehicle's passenger-carrying capacity.

2.1.2 Formulation for the AB-type service

The decision variables/functions for the AB-type design optimization program are listed below,

$s(x)$: stop spacing.

r_c, r_{cc}: the numbers of routes in clockwise and counterclockwise directions, respectively (note the model allows more than two routes in each direction).

$k_c(x)$, $k_{cc}(x)$: the numbers of stops visited by each route in a skip-stop bay in the two directions, respectively;

H_c, H_{cc}: the headways between two consecutive vehicles in the two directions, respectively.

In a given direction, vehicles of each route are dispatched in the same headway, $r_c H_c$ in the clockwise direction and $r_{cc} H_{cc}$ in the counterclockwise direction. The $k_c(x)$ and $k_{cc}(x)$ should take integer values in an AB-type service design. For the simplicity of the solution method, however, we allow these to take any positive real values in the CA model. The optimal $k_c(x)$ and $k_{cc}(x)$ will be converted to integer values in a recipe that generates a real system design from the CA solution, which will be presented in due course.

Next the patrons' travel cost and the agency cost are formulated separately. Patrons' total travel cost consists of total access and egress time, UC_a, total wait time, UC_w (including wait times at both the origin and transfer stops), total in-vehicle travel time, UC_i, and total transfer penalty to account for the inconvenience in transfers, UC_t. First note that the access and egress time UC_a is the same as that of the all-stop service (AUC_a), which was formulated in equation (2.4):

$$UC_a = \int_{x=0}^{L} \frac{s(x)}{4v_w} \big(P_c(x) + Q_c(x) + P_{cc}(x) + Q_{cc}(x)\big) dx \qquad (2.12)$$

The total wait time UC_w is furnished by (2.13) below, in which we define the number of stops in a skip-stop bay that contains x as $T(x) \equiv r_c k_c(x) + 1 = r_{cc} k_{cc}(x) + 1$. These intermediate functions are defined to simplify the model formulation in this section. Note that we specify $r_c k_c(x) = r_{cc} k_{cc}(x)$ for all $x \in (0, L]$ to ensure that the number of stops in a skip-stop bay is the same for both service directions; see constraint (2.22c) later in this section. The $b_c(x)$ and $b_{cc}(x)$ in (2.13) denote the average densities of clockwise and counterclockwise trips that involve backtracking (see again Figure 1.1a) in a skip-stop bay that contains x, and are approximated by (2.14a) and (2.14b), respectively. All these equations are derived in Appendix B.1.

$$UC_w = \frac{(2r_c-1)H_c\Lambda_c}{2} + \frac{(2r_{cc}-1)H_{cc}\Lambda_{cc}}{2} + \int_{x=0}^{L}\left(-\frac{(r_c-1)H_c}{2}\frac{P_c(x)+Q_c(x)}{T(x)} - \frac{(r_{cc}-1)H_{cc}}{2}\frac{P_{cc}(x)+Q_{cc}(x)}{T(x)} + \frac{r_{cc}H_{cc}-r_cH_c}{2}\big(b_c(x)-b_{cc}(x)\big)\right)dx \quad (2.13)$$

$$b_c(x) \approx \frac{r_c(r_c-1)k_c^2(x)}{T^3(x)s(x)}\int_{z\in U^L\left(x,\frac{T(x)s(x)}{2}\right)}\int_{y\in U^L\left(x,\frac{T(x)s(x)}{2}\right)\cap U_+^L(z,T(x)s(x))}\lambda(z,y)dzdy \quad (2.14a)$$

$$b_{cc}(x) \approx \frac{r_{cc}(r_{cc}-1)k_{cc}^2(x)}{T^3(x)s(x)}\int_{z\in U^L\left(x,\frac{T(x)s(x)}{2}\right)}\int_{y\in U^L\left(x,\frac{T(x)s(x)}{2}\right)\cap U_-^L(z,T(x)s(x))}\lambda(z,y)dzdy \quad (2.14b)$$

Equation (2.13) is derived under the assumption that routes are uncoordinated. As shown in Gu et al. (2016), efficient service coordination between these routes at transfer stops will result in much smaller patron wait time. Hence (2.13) is conservative and may overestimate a system's wait time.

Total in-vehicle travel time UC_i is furnished by (2.15) below, where $\eta_c(x)$ and $\eta_{cc}(x)$ are defined in (2.16a-b). These equations are derived in Appendix B.2.

$$UC_i \approx \int_{x=0}^{L}\big(\eta_c(x)+\eta_{cc}(x)\big)dx \quad (2.15)$$

$$\eta_c(x) = o_c(x)\left(\frac{1}{v}+\frac{\tau(k_c(x)+1)}{T(x)s(x)}\right) + b_c(x)\left[\frac{1}{3v}s(x)T(x) + \frac{1}{6}\tau(k_c(x)+k_{cc}(x)+2)\right] \quad (2.16a)$$

$$\eta_{cc}(x) = o_{cc}(x)\left(\frac{1}{v}+\frac{\tau(k_{cc}(x)+1)}{T(x)s(x)}\right) + b_{cc}(x)\left[\frac{1}{3v}s(x)T(x) + \frac{1}{6}\tau(k_c(x)+k_{cc}(x)+2)\right] \quad (2.16b)$$

The total transfer penalty UC_t is approximated by (2.17), where C_t denotes the penalty cost per transfer (in the unit of time). The derivation of (2.17) is presented in Appendix B.3.

$$UC_t \approx C_t\left(\frac{r_c-1}{r_c}\Lambda_c + \frac{r_{cc}-1}{r_{cc}}\Lambda_{cc} - \int_{x=0}^{L}\left(\frac{r_c-1}{r_c}\big(P_c(x)+Q_c(x)\big) + \frac{r_{cc}-1}{r_{cc}}\big(P_{cc}(x)+Q_{cc}(x)\big)\right)\frac{(2T(x)-1)}{2T^2(x)}dx\right) \quad (2.17)$$

The agency cost consists of distance-based and time-based operating costs, AC_K and AC_H, respectively, as well as the amortized line and station infrastructure costs, AC_I and AC_S, respectively. These are formulated as follows:

$$AC_K = \frac{\pi_v L}{\mu}\left(\frac{1}{H_c} + \frac{1}{H_{cc}}\right) \tag{2.18}$$

$$AC_H = \frac{\pi_m}{\mu}\int_{x=0}^{L}\left(\frac{1}{H_c}\left(\frac{1}{v} + \frac{\tau(k_c(x)+1)}{T(x)s(x)}\right) + \frac{1}{H_{cc}}\left(\frac{1}{v} + \frac{\tau(k_{cc}(x)+1)}{T(x)s(x)}\right)\right)dx \tag{2.19}$$

$$AC_I = \frac{2\pi_i L}{\mu} \tag{2.20}$$

$$AC_S = \frac{\pi_s}{\mu}\int_{x=0}^{L}\frac{1}{s(x)}dx \tag{2.21}$$

where π_v, π_m, π_i, and π_s have the same definitions as in Section 2.1.1. Equations (2.18), (2.20) and (2.21) are exactly the same as those in the all-stop model, (2.7), (2.9) and (2.10), respectively. Only (2.19) is different from (2.8) in the all-stop model since the commercial speeds of transit vehicles are different from those in an all-stop system.

The generalized cost minimization problem for the AB-type service can thus be formulated as:

$$\min_{r_c, r_{cc}, H_c, H_{cc}, s(x), k_c(x), k_{cc}(x)} SC = UC_a + UC_w + UC_i + UC_t + AC_K + AC_H + AC_I + AC_S \tag{2.22a}$$

subject to:

$$r_c, r_{cc} \in \{1,2,3,4\} \tag{2.22b}$$

$$r_c k_c(x) = r_{cc} k_{cc}(x), \ \forall x \in (0, L] \tag{2.22c}$$

$$H_{min} + \tau \le H_c \le \frac{K}{\max\limits_{0<x\le L}\{o_c(x)+B(x)\}} \tag{2.22d}$$

$$H_{min} + \tau \le H_{cc} \le \frac{K}{\max\limits_{0<x\le L}\{o_{cc}(x)+B(x)\}} \tag{2.22e}$$

$$s(x), k_c(x), k_{cc}(x) > 0 \tag{2.22f}$$

Constraints (2.22b) specify that there are at most four routes in each direction. Constraints (2.22c) ensure that the transit services of the two directions share the same set of transfer stops (this is required to facilitate backtracking trips). The left parts of (2.22d-e) indicate that the operation headway between two

consecutive vehicles should be no less than the minimum required headway plus the dwell time.

The right parts of (2.22d-e) are the vehicle capacity constraints, where $\max_{0<x\leq L}\{o_c(x)+B(x)\}$ and $\max_{0<x\leq L}\{o_{cc}(x)+B(x)\}$ denote the (approximate) maximum on-board patron flows in the clockwise and counterclockwise directions, respectively. These maximum flows consist of two components: the patron flow of direct travel ($o_c(x)$ and $o_{cc}(x)$), and the extra patron flow added by the backtracking trip segments, denoted by $B(x)$. The latter is derived next. Within a skip-stop bay containing x, the total number of clockwise and counterclockwise trips involving backtracking is $T(x)s(x)\big(b_c(x)+b_{cc}(x)\big)$. Approximately half of these trips will transfer at each of the two transfer stops bounding the bay. Hence, the function $B(x)$ will peak at $\frac{T(x)s(x)}{2}\big(b_c(x)+b_{cc}(x)\big)$ at the downstream side of the upstream transfer stop and the upstream side of the downstream transfer stop. The following approximations are thus used:

$$\max_{0<x\leq L}\{o_c(x)+B(x)\}\approx\max_{0<x\leq L}\left\{o_c(x)+\frac{T(x)s(x)}{2}\big(b_c(x)+b_{cc}(x)\big)\right\} \tag{2.23a}$$

$$\max_{0<x\leq L}\{o_{cc}(x)+B(x)\}\approx\max_{0<x\leq L}\left\{o_{cc}(x)+\frac{T(x)s(x)}{2}\big(b_c(x)+b_{cc}(x)\big)\right\} \tag{2.23b}$$

The assumption behind (2.22d-e) is that patrons behave *uniformly* when choosing transfer stops. For example, it may be assumed that a patron who needs to transfer will always choose to transfer at the first transfer stop on her journey; or, it may be assumed that a patron who needs to transfer will always choose randomly between all the transfer stops along her journey. The latter assumption is reasonable since the choice of transfer stop will not affect the patron's travel cost. If the patrons do not exhibit uniform behavior in the choice of transfer stops, then it may be that one route carries a higher on-board patron flow than other routes at a specific location x, and (2.22d-e) cannot guarantee that any capacity constraint is not violated. For the simplicity of the modeling work, however, this complexity is not considered in this dissertation.

2.2 Solution Method

Sections 2.2.1 and 2.2.2 present solution approaches to the all-stop and AB-type service models, respectively. Section 2.2.3 provides a recipe for converting the optimal CA model solution to a real AB-type service design.

2.2.1 Solution to the all-stop design problem

The all-stop service optimization problem (2.11a-d) has only one decision function $s(x)$ and two scalar decision variables H_c and H_{cc}. This problem can be solved via calculus of variations since the part of the objective function (2.11a) that is related to $s(x)$ is separable by x. This solution method is similar to that described in Wirasinghe and Ghoneim (1981). It is briefly described here to construct the basis for presenting the solution approach to the AB-type service problem in Section 2.2.2.

Inspection of the cost components (2.4-2.10) reveals that the objective function (2.11a), which is the sum of (2.4-2.10), can be expressed as the sum of two parts: one part is only related to the scalar decision variables H_c and H_{cc}, and the other part is an integral over $x \in (0, L]$, whose integrand is related to $s(x)$. These details are as follows:

$$AC = h_A(H_c, H_{cc}) + \int_{x=0}^{L} G_A(H_c, H_{cc}, s(x), x)dx \tag{2.24}$$

where,

$$h_A(H_c, H_{cc}) \equiv \frac{H_c}{2}\Lambda_c + \frac{H_{cc}}{2}\Lambda_{cc} + \frac{\pi_v L}{\mu}\left(\frac{1}{H_c} + \frac{1}{H_{cc}}\right) + \frac{2\pi_i L}{\mu} \tag{2.25a}$$

$$G_A(H_c, H_{cc}, s(x), x) \equiv \frac{s(x)}{4v_w}\left(P_c(x) + Q_c(x) + P_{cc}(x) + Q_{cc}(x)\right) + \left(o_c(x) + o_{cc}(x)\right)\left(\frac{1}{v} + \frac{\tau}{s(x)}\right) + \frac{\pi_m}{\mu}\left(\frac{1}{H_c} + \frac{1}{H_{cc}}\right)\left(\frac{1}{v} + \frac{\tau}{s(x)}\right) + \frac{\pi_s}{\mu s(x)} \tag{2.25b}$$

Note that h_A consists of the total patrons' wait cost (2.5), the distance-based operating cost (2.7), and the line infrastructure cost (2.9). The G_A consists of the *localized* access/egress cost, in-vehicle travel cost, time-based operating cost, and stop infrastructure cost, which are the integrands of (2.4), (2.6), (2.8), and (2.10), respectively.

Since h_A is unrelated to $s(x)$, optimizing (2.24) via an iterative process consists of two stages. In the first stage, optimize $\int_{x=0}^{L} G_A(H_c, H_{cc}, s(x), x)dx$ with respect to $s(x)$ when H_c and H_{cc} are given. The result is the optimal stop spacing function $s^*(x)$ $(0 < x \leq L)$.In the second stage, fix $s(x) = s^*(x)$ for $x \in (0, L]$, and optimize AC with respect to the scalar variables H_c and H_{cc}. The two stages will be iterated until the solutions of $s(x)$, H_c and H_{cc} all converge.

In the first stage, note that the integral $\int_{x=0}^{L} G_A(H_c, H_{cc}, s(x), x)dx$ is separable by x, therefore, minimizing it for the given values of H_c and H_{cc} is equivalent to minimizing $G_A(H_c, H_{cc}, s(x), x)$ for every x. Using the first order condition of (2.25b) with respect to $s(x)$, we find the optimal stop spacing, $s^*(x)$, to minimize $G_A(H_c, H_{cc}, s(x), x)$ for given values of H_c and H_{cc}, is as follows:

$$s^*(x) = \sqrt{\frac{4v_w\left[\tau\left(o_c(x)+o_{cc}(x)+\frac{\pi_m}{\mu}\left(\frac{1}{H_c}+\frac{1}{H_{cc}}\right)\right)+\frac{\pi_s}{\mu}\right]}{P_c(x)+Q_c(x)+P_{cc}(x)+Q_{cc}(x)}} \tag{2.26}$$

In the second stage, we fix $s(x) = s^*(x)$ for $x \in (0, L]$, then the objective function can be rearranged as:

$$AC = \left\{\frac{2\pi_i L}{\mu} + \int_{x=0}^{L}\left(\frac{s^*(x)}{4v_w}\left(P_c(x) + Q_c(x) + P_{cc}(x) + Q_{cc}(x)\right) + \left(o_c(x) + o_{cc}(x)\right)\left(\frac{1}{v} + \frac{\tau}{s^*(x)}\right) + \frac{\pi_s}{\mu s^*(x)}\right)dx\right\} + \left\{\frac{H_c}{2}\Lambda_c + \frac{H_{cc}}{2}\Lambda_{cc} + \frac{\pi_v L}{\mu}\left(\frac{1}{H_c} + \frac{1}{H_{cc}}\right) + \int_{x=0}^{L}\frac{\pi_m}{\mu}\left(\frac{1}{v} + \frac{\tau}{s^*(x)}\right)\left(\frac{1}{H_c} + \frac{1}{H_{cc}}\right)dx\right\} \tag{2.27}$$

The cost terms in the first pair of braces in (2.27) are unrelated toH_c and H_{cc}, while the expression in the second pair of braces can be minimized by taking the first order conditions with respect to H_c and H_{cc}, respectively:

$$\widetilde{H}_c = \sqrt{\frac{2}{\Lambda_c\mu}\left(\pi_v L + \frac{\pi_m L}{v} + \int_{x=0}^{L}\frac{\pi_m\tau}{s^*(x)}dx\right)} \tag{2.28a}$$

$$\widetilde{H}_{cc} = \sqrt{\frac{2}{\Lambda_{cc}\mu}\left(\pi_v L + \frac{\pi_m L}{v} + \int_{x=0}^{L}\frac{\pi_m\tau}{s^*(x)}dx\right)} \tag{2.28b}$$

To calculate the integrals in (2.28a-b), the range $(0, L]$ is discretized into many equal-sized small intervals, and the values of $s^*(x)$ are calculated using (2.26) only for x values at the midpoint of each interval. Then the integrals are approximated using the trapezoidal rule.

Considering constraints (2.11b-c), we have:

$$H_c^* = \text{mid}\left\{H_{min}, \widetilde{H}_c, \frac{K}{\max\limits_{0<x\le L}\{o_c(x)\}}\right\} \tag{2.29a}$$

$$H_{cc}^* = \text{mid}\left\{H_{min}, \widetilde{H}_{cc}, \frac{K}{\max\limits_{0<x\le L}\{o_{cc}(x)\}}\right\} \tag{2.29b}$$

where function $\text{mid}\{x, y, z\}$ returns the middle value among x, y and z (Daganzo, 2007).

With the updated H_c and H_{cc} calculated by (2.29a-b), return to the first stage to update $s(x)$. The two stages are iterated alternately until the solution converges. The algorithm is thus summarized as:

Step 1. Discretize the continuous range of x, $(0, L]$, into n equal-sized intervals, each of length $\Delta x \equiv \frac{L}{n}$. Write $x_j \equiv \left(j - \frac{1}{2}\right)\Delta x$ for $(j = 1,2, \dots, n)$. Initialize H_c and H_{cc} satisfying $H_{min} \le H_c, H_{cc} \le \frac{K}{\max\limits_{0<x\le L}\{o_c(x)\}}$.

Step 2. For each $j \in \{1,2, \dots, n\}$, apply (2.26) to calculate $s^*(x_j)$.

Step 3. Apply (2.28a-29b) to calculate H_c^* and H_{cc}^*. The integrals in (2.28a-b) are calculated using the trapezoidal rule.

Step 4. Set $H_c = H_c^*$, $H_{cc} = H_{cc}^*$, and $s(x_j) = s^*(x_j)$ for $j = 1,2, \dots, n$. Repeat Steps 2 and 3. If $|H_c - H_c^*| + |H_{cc} - H_{cc}^*| \le \epsilon_1$ and $\sum_{j=1}^{n}|s(x_j) - s^*(x_j)| \le \epsilon_2$ are both satisfied for prespecified tolerance values ϵ_1 and ϵ_2, then consider the solution converged. Otherwise, repeat Step 4.

The all-stop program (2.11a-d) is convex, and the above algorithm will converge to the global optimum if ϵ_1 and ϵ_2 approach to zero. Hence, the initial values of H_c and H_{cc} will not affect this convergence and the algorithm's solution quality. The detailed proofs of the convexity of (2.11a-d) and the solution algorithm's convergence are omitted here because the all-stop model is not the focus of this dissertation.

2.2.2 Solution to the AB-type design problem

The solution procedure to the AB-type service problem is similar to that of the all-stop problem presented above. The objective function is also decomposed (2.22a)

into two parts: one part as denoted by $h_S(\cdot)$ is related to the scalar decision variables r_c, r_{cc}, H_c and H_{cc}; and the other part as denoted by $\int_{x=0}^{L} G_S(\cdot,x)dx$, is related to the decision functions $s(x)$, $k_c(x)$ and $k_{cc}(x)$. We then optimize $SC \equiv h_S(\cdot) + \int_{x=0}^{L} G_S(\cdot,x)dx$ again by alternately performing two stages: i) fix all the scalar decision variables and find the optimal decision functions that minimize $G_S(\cdot,x)$ for each $x \in (0,L]$; and ii) fix the decision function values and find the optimal scalar variable values that minimize SC.

Specifically, first fix r_c and r_{cc} (because they take values from a small finite set $\{1,2,3,4\}$), and replace $k_c(x)$ and $k_{cc}(x)$ by $\frac{T(x)-1}{r_c}$ and $\frac{T(x)-1}{r_{cc}}$, respectively (see the definition of $T(x)$ in Section 2.1.2). Also fix $b_c(x)$ and $b_{cc}(x)$ so that the resulting objective function is separable by x. Then by inspecting the cost components (2.12-2.21), rewrite the objective function (2.22a) for given r_c and r_{cc} as follows:

$$SC|r_c, r_{cc} = h_S(H_c, H_{cc}) + \int_{x=0}^{L} G_S(H_c, H_{cc}, s(x), T(x), x)dx \tag{2.30}$$

where,

$$h_S(H_c, H_{cc}) \equiv \frac{(2r_c-1)\Lambda_c}{2} H_c + \frac{(2r_{cc}-1)\Lambda_{cc}}{2} H_{cc} + C_t\left(\frac{r_c-1}{r_c}\Lambda_c + \frac{r_{cc}-1}{r_{cc}}\Lambda_{cc}\right) + \frac{\pi_v L}{\mu}\left(\frac{1}{H_c} + \frac{1}{H_{cc}}\right) + \frac{2\pi_i L}{\mu} \tag{2.31a}$$

$$G_S(H_c, H_{cc}, s(x), T(x), x) \equiv \frac{P_c(x)+Q_c(x)+P_{cc}(x)+Q_{cc}(x)}{4v_w} s(x) - \frac{(r_c-1)\left(P_c(x)+Q_c(x)\right)}{2T(x)} H_c - \frac{(r_{cc}-1)\left(P_{cc}(x)+Q_{cc}(x)\right)}{2T(x)} H_{cc} - \frac{r_c\left(b_c(x)-b_{cc}(x)\right)}{2} H_c + \frac{r_{cc}\left(b_c(x)-b_{cc}(x)\right)}{2} H_{cc} + \frac{o_c(x)+o_{cc}(x)}{v} + \frac{o_c(x)\tau}{r_c}\frac{T(x)+r_c-1}{T(x)s(x)} + \frac{o_{cc}(x)\tau}{r_{cc}}\frac{T(x)+r_{cc}-1}{T(x)s(x)} + \frac{b_c(x)+b_{cc}(x)}{3v} s(x)T(x) + \frac{1}{6}\tau\left(b_c(x)+b_{cc}(x)\right)\left(\frac{T(x)-1}{r_c} + \frac{T(x)-1}{r_{cc}} + 2\right) - C_t\left(\frac{r_c-1}{r_c}\left(P_c(x)+Q_c(x)\right) + \frac{r_{cc}-1}{r_{cc}}\left(P_{cc}(x)+Q_{cc}(x)\right)\right)\frac{(2T(x)-1)}{2T^2(x)} + \frac{\pi_m}{\mu}\left(\frac{H_c+H_{cc}}{v} + \frac{\tau}{r_c}\frac{T(x)+r_c-1}{H_c T(x)s(x)} + \frac{\tau}{r_{cc}}\frac{T(x)+r_{cc}-1}{H_{cc}T(x)s(x)}\right) + \frac{\pi_s}{\mu} s^{-1}(x) \tag{2.31b}$$

The h_S consists of part of the total patrons' wait cost (2.13), part of the total transfer penalty (2.17), the distance-based operating cost (2.18), and the line infrastructure cost (2.20). The G_S consists of the localized access/egress cost (2.12),

wait cost in (2.13), in-vehicle travel cost in (2.15), time-based operating cost (2.19), and stop infrastructure cost (2.21).

In the first stage, the optimal values of $s(x)$ and $T(x)$ for minimizing $G_S(H_c, H_{cc}, s(x), T(x), x)$ can be obtained when H_c, H_{cc}, and x are given. Note that the formulas of $b_c(x)$ and $b_{cc}(x)$ (2.14a-b) contain double integrals. Thus fixing $b_c(x)$ and $b_{cc}(x)$ is necessary to ensure $G_S(\cdot, x)$ is separable by x and $s(x)$ and $T(x)$ are the only two decision variables to be optimized at this stage for any given x. Note too that when $b_c(x)$ and $b_{cc}(x)$ are fixed, $G_S(\cdot, x)$ is convex with respect to $s(x)$. This can be verified by checking that the second order derivative of $G_S(\cdot, x)$ with respect to $s(x)$ is positive. Hence, the (unconstrained) optimal stop spacing, $\tilde{s}(x)$ can be derived from the first order condition of (2.31b) as follows:

$$\tilde{s}(x)|H_c, H_{cc}, b_c(x), b_{cc}(x) = \sqrt{\frac{\left(\frac{1}{r_c}+\frac{r_c-1}{r_c T(x)}\right)\left(o_c(x)+\frac{\pi_m}{\mu H_c}\right)\tau+\left(\frac{1}{r_{cc}}+\frac{r_{cc}-1}{r_{cc} T(x)}\right)\left(o_{cc}(x)+\frac{\pi_m}{\mu H_{cc}}\right)\tau+\frac{\pi_s}{\mu}}{\left(P_c(x)+Q_c(x)+P_{cc}(x)+Q_{cc}(x)\right)/4v_w+T(x)\left(b_c(x)+b_{cc}(x)\right)/3v}} \quad (2.32)$$

Note that the right parts of constraints (2.22d-e) can be re-written as:

$$s(x) \le \frac{\min\{2K/H_c-2o_c(x), 2K/H_{cc}-2o_{cc}(x)\}}{\left(b_c(x)+b_{cc}(x)\right)T(x)}.$$

Hence, the optimal $s(x)$ is given by:

$$s^*(x)|H_c, H_{cc}, b_c(x), b_{cc}(x) = \min\left\{\tilde{s}(x), \frac{\min\{2K/H_c-2o_c(x), 2K/H_{cc}-2o_{cc}(x)\}}{\left(b_c(x)+b_{cc}(x)\right)T(x)}\right\} \quad (2.33)$$

By plugging (2.33) into (2.31b), the G_S is converted to a function of $T(x)$ only (while $H_c, H_{cc}, b_c(x)$ and $b_{cc}(x)$) are still given. The minimization of G_S with respect to $T(x)$ can be quickly solved to near optimality by a number of methods, including the exhaustive search method. To ensure solution quality, the exhaustive search is used in this paper, since it guarantees that the solution $(s^*(x), T^*(x)|H_c, H_{cc}, b_c(x), b_{cc}(x))$ converges to the global optimum when the search interval approaches zero. This is because the objective function is continuous and has bounded derivatives with respect to $T(x)$. Note again that in the CA model we allow $T(x)$ to take positive real values, and it will be converted to integers in the recipe presented in Section 2.2.3.

After calculating $s^*(x)$ and $T^*(x)$ for given $H_c, H_{cc}, b_c(x)$ and $b_{cc}(x)$, update $b_c(x)$ and $b_{cc}(x)$ using $s^*(x)$ and $T^*(x)$. The G_S will then be re-optimized with the updated $b_c(x)$ and $b_{cc}(x)$. The process is iterated until $s^*(x)$, $T^*(x)$, $b_c(x)$ and $b_{cc}(x)$ all converge. At the end of this first stage, $s^*(x)$ and $T^*(x)$ for the given H_c and H_{cc} are obtained.

In the second stage, fix $s(x)$ and $T(x)$ to the optimal values obtained in the first stage for $x \in (0, L]$, and optimize SC with respect to the scalar variables H_c and H_{cc} (while r_c and r_{cc} are still kept constant for now). By taking first-order derivatives of SC with respect to H_c and H_{cc}, the unconstrained optima of H_c and H_{cc} are obtained as follows:

$$\widetilde{H}_c = \sqrt{\frac{\frac{\pi_v L}{\mu}+\frac{\pi_m}{\mu}\int_{x=0}^{L}\left(\frac{1}{v}+\frac{\tau}{s^*(x)}\left(\frac{1}{r_c}+\frac{r_c-1}{r_c T^*(x)}\right)\right)dx}{\frac{2r_c-1}{2}\Lambda_c+\int_{x=0}^{L}\left(-\frac{r_c-1}{2}\frac{P_c(x)+Q_c(x)}{T^*(x)}-\frac{r_c}{2}b_c(x)+\frac{r_c}{2}b_{cc}(x)\right)dx}} \tag{2.34a}$$

$$\widetilde{H}_{cc} = \sqrt{\frac{\frac{\pi_v L}{\mu}+\frac{\pi_m}{\mu}\int_{x=0}^{L}\left(\frac{1}{v}+\frac{\tau}{s^*(x)}\left(\frac{1}{r_{cc}}+\frac{r_{cc}-1}{r_{cc}T^*(x)}\right)\right)dx}{\frac{2r_{cc}-1}{2}\Lambda_{cc}+\int_{x=0}^{L}\left(-\frac{r_{cc}-1}{2}\frac{P_{cc}(x)+Q_{cc}(x)}{T^*(x)}-\frac{r_{cc}}{2}b_{cc}(x)+\frac{r_{cc}}{2}b_c(x)\right)dx}} \tag{2.34b}$$

By considering the boundary constraints for the headways (2.22d-e), the optimal headways are obtained as follows:

$$H_c^* = \text{mid}\left\{H_{min}, \widetilde{H}_c, \frac{K}{\max\limits_{0<x\le L}\{o_c(x)+B(x)\}}\right\} \tag{2.35a}$$

$$H_{cc}^* = \text{mid}\left\{H_{min}, \widetilde{H}_{cc}, \frac{K}{\max\limits_{0<x\le L}\{o_{cc}(x)+B(x)\}}\right\} \tag{2.35b}$$

where function $\text{mid}\{x, y, z\}$ returns the middle value among x, y and z (Daganzo, 2007).

The two stages described above will be iterated alternately until the solution converges for given r_c and r_{cc}. This process is then repeated for all the possible combinations of $(r_c, r_{cc}) \in \{1,2,3,4\} \times \{1,2,3,4\}$ (see constraint (2.22b)), where "×" represents the Cartesian product operator. The lowest-cost solution is then identified.

The entire solution algorithm is summarized as follows:

Step 1. Discretize the continuous range of x, $(0, L]$, into n_1 equal-sized intervals, each of length $\Delta x \equiv \frac{L}{n_1}$. Write $x_j \equiv \left(j - \frac{1}{2}\right)\Delta x$ for $(j = 1,2, \ldots, n)$.

Step 2. Select an unvisited pair of (r_c, r_{cc}) from $\{1,2,3,4\} \times \{1,2,3,4\}$.

Step 3. Initialize H_c and H_{cc} satisfying $H_{min} \leq H_c \leq \frac{K}{\max\limits_{0<x\leq L}\{o_c(x)\}}$, and $H_{min} \leq H_{cc} \leq \frac{K}{\max\limits_{0<x\leq L}\{o_{cc}(x)\}}$.

Step 4. For each $x = x_j, j \in \{1,2, \ldots, n_1\}$:

Step 4.1. Initialize $b_c(x) = b_{cc}(x) = 0$.

Step 4.2. Select an unvisited value of $T(x)$ from a finite set, Ψ, which contains the discrete values of $T(x)$ that will be searched exhaustively. Calculate $s^*(x)$ using (2.32-33) and the optimal function value of $G_S\left(s^*(x), x|T(x), H_c, H_{cc}, b_c(x), b_{cc}(x)\right)$.

Step 4.3. Repeat step 4.2 until Ψ is exhausted. Record the $T^*(x)$ as the $T(x)$ value that minimizes G_S.

Step 4.4. Calculate a new pair of values for $b_c(x)$ and $b_{cc}(x)$ using (15a-b) and $s^*(x)$, $T^*(x)$; denote them as $\hat{b}_c(x)$ and $\hat{b}_{cc}(x)$. Update $b_c(x)$ and $b_{cc}(x)$ as follows: $b_c(x) \leftarrow (1 - \alpha_1)b_c(x) + \alpha\hat{b}_c(x)$ and $b_{cc}(x) \leftarrow (1 - \alpha)b_{cc}(x) + \alpha\hat{b}_{cc}(x)$, where α satisfies $0 < \alpha < 1$.

Step 4.5. Set $\hat{s}(x) = s^*(x)$, $\hat{T}(x) = T^*(x)$. Repeat steps 4.2-4.4 until $|s^*(x) - \hat{s}(x)| + \left|T^*(x) - \hat{T}(x)\right| \leq \epsilon_3$ and $\left|b_c(x) - \hat{b}_c(x)\right| + \left|b_{cc}(x) - \hat{b}_{cc}(x)\right| \leq \epsilon_4$ are both satisfied for prespecified tolerance values ϵ_3 and ϵ_4.

Step 5. Apply (2.34a-2.35b) to calculate H_c^* and H_{cc}^*. The integrals in (34a-b) are calculated using the trapezoidal rule. Update H_c and H_{cc} as follows: $H_c \leftarrow H_c^*$ and $H_{cc} \leftarrow H_{cc}^*$,

Step 6. Set $\hat{s}(x) = s^*(x)$, $\hat{T}(x) = T^*(x)$. Repeat steps 4-5 until $|s^*(x) - \hat{s}(x)| + \left|T^*(x) - \hat{T}(x)\right| \leq \epsilon_3$ and $|H_c - H_c^*| + |H_{cc} - H_{cc}^*| \leq \epsilon_5$ are both satisfied for prespecified tolerance values ϵ_3 and ϵ_5.

Step 7. Repeat steps 2-6 until the feasible range of (r_c, r_{cc}) is exhausted. Record the solution $\left(r_c^*, r_{cc}^*, H_c^*, H_{cc}^*, s^*(x), T^*(x)\right)$ that yields the lowest value of SC. Set $k_c^*(x) = \frac{T^*(x)-1}{r_c^*}$ and $k_{cc}^*(x) = \frac{T^*(x)-1}{r_{cc}^*}$.

Unfortunately, the program (2.22a-f) is not convex, mainly due to the existence of $b_c(x)$ and $b_{cc}(x)$. Hence, a lower bound of the optimal cost SC has also been developed. The derivation of this lower bound can be found in Appendix C. In our numerical analysis in Section 2.3, we compare our solution obtained from the above algorithm against the lower bound. This solution is found to be very close to the global optimum.

2.2.3 Generating the exact stop locations

The solution obtained using the above algorithm includes arrays of discrete points on the continuous functions $s(x)$, $k_c(x)$, and $k_{cc}(x)$. A 3-step recipe is proposed below to generate a real stop location and route plan. In Step 1 of this recipe, the stop locations are generated from the optimal $s(x)$, using a method similar to Wirasinghe and Ghoneim (1981). In Step 2, the transfer stops are selected from the set of stops generated in Step 1, by using the optimal r_c, r_{cc}, $k_c(x)$ and $k_{cc}(x)$. Note that the continuous solutions of $k_c(x)$ and $k_{cc}(x)$ are non-integer, and in Step 2 it is ensured that they take integer values that satisfy (2.21c) in each skip-stop bay. Step 3 completes the design by assigning the non-transfer stops to each route.

Step 1. Place one stop at every x where $\int_{z=0}^{x} \frac{dz}{s(z)}$ is an integer (the first stop is located at $x = 0$)[4]. If the last stop is located at $\tilde{x}$ and $L - \tilde{x} < \frac{s(L)}{2}$ (i.e. the last stop is too close to the first one), remove that last stop. Denote the resulting stop location plan as $\mathbf{\Omega} = \left\{x_i^S : i = 1, 2, \dots, N^S\right\}$, where N^S denotes the number of stops, and x_i^S the location of the i-th stop satisfying $0 = x_1^S < x_2^S < \cdots < x_{N^S}^S < L$.

[4] For a linear corridor, stops should be located where $\int_{z=0}^{x} \frac{dz}{s(z)} - \frac{1}{2}$ is an integer because the first stop is best located half a stop spacing from $x = 0$.

Step 2. Select the first stop ($x_1^S = 0$) to be a transfer stop. Select the remaining transfer stops from $\boldsymbol{\Omega}$ recursively as follows: given the location of the j-th transfer stop ($j \geq 1$), x_j^T, find $x_{j+1}^T \in \boldsymbol{\Omega}$ which minimizes $\left| \int_{z=x_j^T}^{x_{j+1}^T} \frac{dz}{s(z)} - r_c \frac{\int_{z=x_j^T}^{x_{j+1}^T} k_c(z)dz}{x_{j+1}^T - x_j^T} - 1 \right|$, subject to: i) $x_{j+1}^T > x_j^T$; and ii) $\int_{z=x_j^T}^{x_{j+1}^T} \frac{dz}{s(z)} - 1$ is an integer multiple of both r_c and r_{cc}. If the last transfer stop is located at $\tilde{x}^T$ and $\int_{z=\tilde{x}^T}^{L} \frac{dz}{s(z)} < \frac{1}{2}\left(r_c \frac{\int_{z=\tilde{x}^T}^{L} k_c(z)dz}{L - \tilde{x}^T} + 1 \right)$ (i.e., if the last transfer stop is too close to the first one), remove that last transfer stop. Denote N^T as the number of transfer stops selected. Note that in the last skip-stop bay (the one between the last transfer stop at $x_{N^T}^T$ and $x_1^T = 0$), the above constraint ii) may not be satisfied. This may result in some routes visiting one less stop than the other routes in the same skip-stop bay.

Step 3. Between any two consecutive transfer stops, the non-transfer stops are assigned to the r_c (or r_{cc}) routes in rotation.

After generating the real stop location and route plan, the generalized cost for the AB-type service is recalculated; see Ulusoy et al. (2010) and Leiva et al. (2010) for similar cost models that can be used for this purpose.

2.3 Numerical Analysis

Section 2.3.1 describes the demand patterns and other parameter values used in our numerical experiments. Section 2.3.2 reports the validation tests of our CA models. Section 2.3.3 examines the optimal design of the AB-type service under a symmetric demand pattern. Section 2.3.4 presents the parametric analysis of the optimal design, also under symmetric demand patterns. Section 2.3.5 provides examples under asymmetric demand patterns.

2.3.1 Demand patterns and parameter values

We consider a demand density function of the following form:

$$\lambda(x,y)=\begin{cases} p_c(x)\theta_c\big(l(x,y)\big)\Lambda_c, & \text{if } 0<y-x\le\frac{L}{2} \text{ or } y-x\le-\frac{L}{2} \\ p_{cc}(x)\theta_{cc}\big(l(x,y)\big)\Lambda_{cc}, & \text{if } -\frac{L}{2}<y-x\le 0 \text{ or } y-x>\frac{L}{2} \end{cases} \tag{2.36}$$

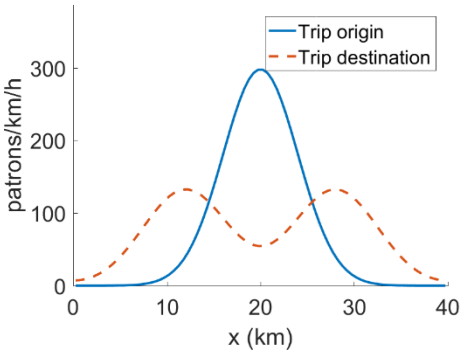
where $p_c(x)$ and $p_{cc}(x)$ are the probability density functions (PDFs) of the trip origins over the corridor for clockwise and counterclockwise directions, respectively; $\theta_c(l)$ and $\theta_{cc}(l)$ are the PDFs of trip length, where $l(x,y)=\min(|x-y|,L-|x-y|)$ denotes the trip length, $0<l\le\frac{L}{2}$. This form of demand function allows us to examine separately the effects on the optimal transit corridor design stemming from: i) the spatial heterogeneity in trip origins; and ii) the distribution of trip lengths.

In the numerical experiments presented in the following sections, $p_c(x)$ and $p_{cc}(x)$ are assumed to be truncated normal PDFs denoted by $Tr\mathcal{N}\left(\frac{L}{2},\sigma_o^2,0,L\right)$; i.e., normal distributions with mean $\frac{L}{2}$ and variance σ_o^2, truncated by the interval $[0,L]$. The standard deviation parameter σ_o can take different values between the clockwise and counterclockwise directions. The $\theta_c(l)$ and $\theta_{cc}(l)$ are assumed to be PDFs of uniform distributions denoted by $\mathcal{U}(E_l-\sqrt{3}\sigma_l,E_l+\sqrt{3}\sigma_l)$, where E_l and σ_l are the mean and standard deviation of the trip length, respectively (these can also take different values between the two service directions). We further specify that $\sigma_o\in\{\infty,8,4\}$ (km); the three values indicate no spatial variation (i.e. uniform distribution), low spatial variation and high spatial variation, respectively, for trip origin distribution. Also, four trip length distributions with $(E_l,\sigma_l)\in\{(8,2),(8,4),(12,2),(12,4)\}$ (km, km) are examined. The curves of trip origin and destination densities, $P_c(x)+P_{cc}(x)$ and $Q_c(x)+Q_{cc}(x)$, for the eight heterogeneous demand patterns in a 40-km corridor are plotted in Figure 2.2a-h: they are plotted for $\sigma_o=4$ and $(E_l,\sigma_l)\in\{(8,2),(8,4),(12,2),(12,4)\}$ in Figure 2.2a-d, and for $\sigma_o=8$ and the same four trip length distributions in Figure 2.2e-h. Each plot assumes that the demand is *symmetric* in both directions with average density $\frac{\Lambda_c}{L}=\frac{\Lambda_{cc}}{L}=37.5$ trips/km/h. These figures illustrate how demand pattern changes along with the key parameters defined above.

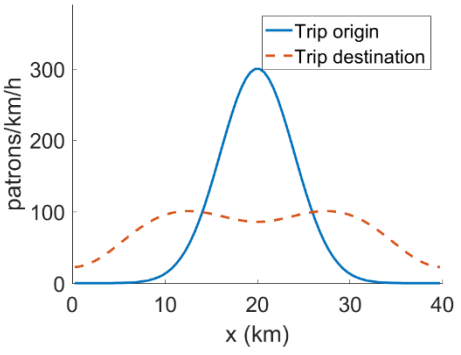
The following sections examine two typical transit modes, bus and rail, operating on a 40-km corridor with cost and operational parameters borrowed from

previous studies (Daganzo, 2010; Sivakumaran et al., 2014; Gu et al., 2016). These parameter values are listed in Table 2.1. We examine three levels of average demand densities for each transit mode: $\frac{\Lambda_c}{L}, \frac{\Lambda_{cc}}{L} \in \{37.5, 75, 150\}$ trips/km/h for a bus corridor, and $\{250, 500, 1000\}$ trips/km/h for a rail corridor; and two values of time: $\mu = 5$ \$/h for a low-wage city, and $\mu = 20$ \$/h for a high-wage city. A low walking speed ($v_w = 2$ km/h) is assumed to account for the signal delays and the inconvenience of walking; C_t is assumed to be 1 min/transfer; and α_1 and α_2 are set to be 0.5 in the solution algorithm.

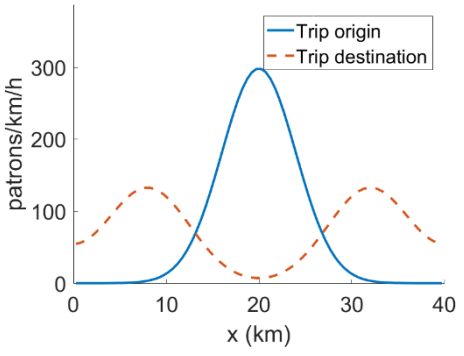
For our solution algorithm, the decompose interval Δx is 0.5 km, the smooth factor $\alpha = 0.5$, and the convergence tolerance $\epsilon_1 = \epsilon_2 = \epsilon_3 = \epsilon_4 = \epsilon_5 = 0.001$.



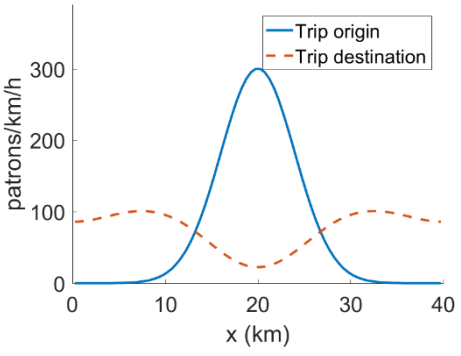
(a) Highly-spatially-varied demand with short, less-varied trip lengths ($E_l = 8, \sigma_l = 2, \sigma_o = 4$ km).



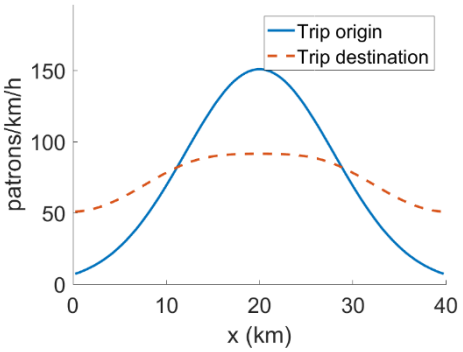
(b) Highly-spatially-varied demand with short, highly-varied trip lengths ($E_l = 8, \sigma_l = 4, \sigma_o = 4$ km).



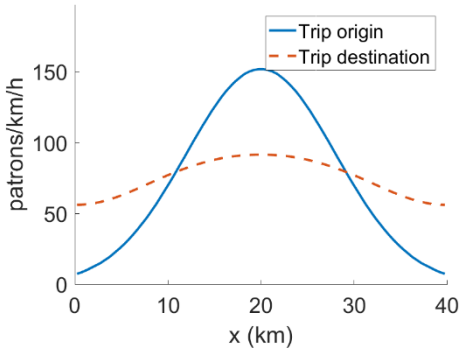
(c) Highly-spatially-varied demand with long, less-varied trip lengths ($E_l = 12, \sigma_l = 2, \sigma_o = 4$ km).



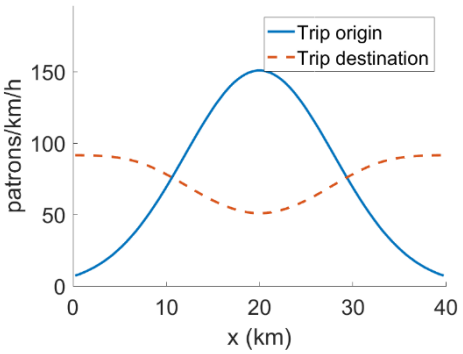
(d) Highly-spatially-varied demand with long, highly-varied trip lengths ($E_l = 12, \sigma_l = 4, \sigma_o = 4$ km).



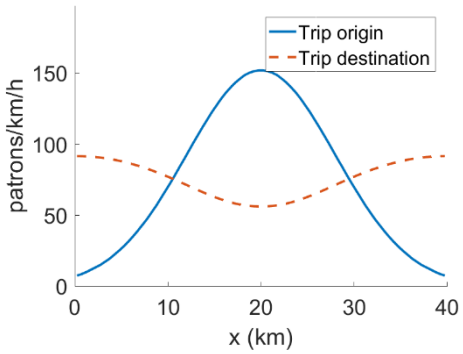
(e) Less-spatially-varied demand with short, less-varied trip lengths ($E_l = 8, \sigma_l = 2, \sigma_o = 8$ km).



(f) Less-spatially-varied demand with short, highly-varied trip lengths ($E_l = 8, \sigma_l = 4, \sigma_o = 8$ km).



(g) Less-spatially-varied demand with long, less-varied trip lengths ($E_l = 12, \sigma_l = 2, \sigma_o = 8$ km).



(h) Less-spatially-varied demand with long, highly-varied trip lengths ($E_l = 12, \sigma_l = 4, \sigma_o = 8$ km).

Figure 2.2 Trip origin and destination densities along the corridor.

Table 2.1 Cost and operational parameters for bus and rail

	π_v (\$/veh·km)	π_m (\$/veh·h)	π_i (\$/km/h)	π_s (\$/stop/h)
Bus	0.59	$2.66 + 3\mu$	$6 + 0.2\mu$	$0.42 + 0.014\mu$
Rail	2.20	$101 + 5\mu$	$594 + 19.8\mu$	$294 + 9.8\mu$
	τ (sec)	v (km/h)	K (passenger/veh)	H_{min} (min)
Bus	30	25	80	1
Rail	45	60	3000	1.5

2.3.2 Model validation

Since the all-stop model is a special case of the AB-type service model with $r_c = r_{cc} = 1$, this section examines the solution quality and the computational efficiency for the AB-type service model only. Specifically, solution quality testing is presented in Section 2.3.2.1; computational efficiency is examined in

Section 2.3.2.2. How the discretization interval Δx (see Step 1 in Section 2.2.2) affects the solution quality and computation time is examined in Section 2.3.2.3.

2.3.2.1 Solution quality

We first examine the quality of our CA solution to the AB-type service model (2.22a-f), and then investigate the approximation error of the CA model as compared against the converted real design (see Section 2.2.3).

The exact solution to (2.22a-f) is very difficult to obtain due to the non-convexity of the program. Hence, the lower bound developed in Appendix C is used to evaluate the solution quality obtained by the algorithm presented in Section 2.2.2. The lower bound is developed by replacing the original objective function (2.22a) with a lower bound of (2.22a), and relaxing some original constraints. The resulting program's exact solution can be found via the method described in Appendix C.

The relative cost gap between the lower bound and the CA solution obtained by the algorithm in Section 2.2.2 is calculated for all the 144 symmetric-demand instances (see Section 2.3.1). The average cost gap is found to be only 0.7% and the maximum cost gap is 2.8%. This shows that the solutions found by our algorithm are quite good.[5]

Approximation errors between the CA solution and the converted real design for the 144 numerical instances are summarized in Table 2.2. Note first that the generalized cost error never exceeds 1.2% and averages only 0.2%. Thus, our CA model's generalized cost estimation is quite accurate.

This high level of accuracy can be partly explained as follows. Our CA model's generalized cost function (like many other CA formulations for transit corridor or network design problems in the literature) exhibits a form similar to a generalized Economic Order Quantity model $Ay^a + By^{-b}$, where y is the

[5] Note that the lower bound is developed mainly by ignoring the costs related to backtracking. Hence the modest cost gaps reported above imply that the backtracking-related cost is very small (compared to other cost components). Note, however, that the backtracking costs cannot be ignored in our optimization model, because otherwise the skip-stop bays would be unreasonably large in the optimal design (without considering backtracking, only one transfer stop would be needed throughout the corridor).

decision variable and A, B, a, b are positive parameters. For example, function G_S in (2.32b) has the above form with respect to $s(x)$. For a model of this kind, the objective function value is quite insensitive to small changes in the decision variable(s) near the optimal solution, because the resulting changes in the cost components are in different directions and are largely canceled out (Daganzo, 2010).

Table 2.2 , in addition, shows that most of the detailed cost components also have very small errors. However, larger errors are observed for the transfer cost UC_t (6.1% on average and 38.3% at maximum), which are due to the approximation used in the derivation of equation (2.17) in Appendix B. Moderate errors are also observed for the waiting cost UC_w (only 2.1% on average but 10.4% at maximum). This is caused by estimating the backtracking trip densities $b_c(x)$ and $b_{cc}(x)$ using equations (2.15a-b).

Table 2.2 Comparing the costs of CA solution and the converted real design

	Average error (%)	Maximum error (%)
Generalized cost, SC	0.2%	1.2%
User cost	0.3%	1.5%
Agency cost	0.3%	1.5%
UC_a	0.5%	1.3%
UC_w	2.1%	10.4%
UC_i	0.4%	2.3%
UC_t	6.1%	38.3%
AC_K	0.0%	0.0%
AC_H	0.5%	2.5%
SC_I	0.0%	0.0%
AC_S	0.5%	1.2%

2.3.2.2 Computation time

Computation times for all the 144 symmetric-demand instances are summarized in Table 2.3. The run times for three stages of the solution procedure are reported separately: i) demand preprocessing, which calculates the aggregate demand functions (2.1a-2.3b); ii) finding the CA solution via the algorithm presented in Section 2.2.2; and iii) generating a real design plan via the recipe in Section 2.2.3.

These stages are performed in Matlab 2016a on a personal computer with Intel Core i7-4970 CPU @ 3.60 GHz and 16.00 G RAM.

Table 2.3 shows that the demand preprocessing stage is the most time-consuming, which takes 2-3 minutes in general. The second stage is quite fast, taking only 13.9 seconds on average. This indicates that our solution algorithm is very efficient. The last stage takes only less than 1 second to complete. The total run times show that our solution procedure is fairly efficient.

Table 2.3 Computation time of the solution procedure

	Average run time (s)	Maximum run time (s)
Demand preprocessing	137.3	141.2
Finding the CA solution	13.9	35.5
Generating a real design plan	0.2	0.7
Total	151.4	175.6

2.3.2.3 Choice of the discretization interval Δx

Our solution algorithm requires that we discretize the spatial coordinate of the corridor by an interval Δx. A larger Δx will increase the error due to the discretization, while a smaller Δx would increase the run time. This is illustrated in Table 2.4, where the run times and optimal generalized cost changes obtained from the CA model are summarized for $\Delta x = 0.25, 0.5, 1, 2$, and 4 km. Note that the solution procedure's total run time, averaged across all the 144 numerical instances, declines rapidly as Δx increases (in fact the average run time is approximately proportional to $1/\Delta x$).On the other hand, the average percentage change in the CA model's optimal generalized cost (using the case of $\Delta x = 0.25$ km as a benchmark for calculating the percentage change) increases very slowly with Δx, especially when $\Delta x = 0.25, 0.5$, and 1 km. The maximum percentage change in the optimal generalized cost among all the 144 instances is only 0.6% for $\Delta x = 0.25$ and 0.5 km. The table also provides the percentage errors between the CA solution's generalized costs and the converted real design. These percentage values show that the error increases with Δx, but is never greater than 1.3% for $\Delta x = 0.25, 0.5$, and 1 km. These results go some way to verifying our solution procedure's robustness and the appropriateness of choosing $\Delta x = 0.5$ km.

Table 2.4 Comparison between different discretization intervals

Δx (km)	0.25	0.5	1	2	4
Average total run time (s)	413.7	151.4	71.3	35.8	16.3
Average percentage change in the generalized cost of CA model	-	0.4%	0.5%	0.6%	1.7%
Maximum percentage change in the generalized cost of CA model	-	0.6%	1.1%	1.8%	7.7%
Average percentage error between the costs of CA model and the converted design	0.2%	0.2%	0.3%	0.4%	1.1%
Maximum percentage error between the costs of CA model and the converted design	0.7%	1.2%	1.3%	2.5%	5.4%

2.3.3 Optimal design of AB-type service

In this section, we examine the optimal design of an AB-type service bus corridor for a high-wage city ($\mu = 20$ \$/h) with high spatially-varied origin densities ($\sigma_o = 4$ km), long and less-varied trip lengths ($E_l = 12$ km, $\sigma_l = 2$ km), and symmetric demand pattern with low density ($\frac{\Lambda_c}{L} = \frac{\Lambda_{cc}}{L} = 37.5$ trips/km/h). The optimal design features two routes in each direction. The converted stop location plan is presented in Figure 2.3a, where routes 1 and 2's non-transfer stops are marked by blue rings and red squares, respectively, and transfer stops are marked by solid black circles. The figure shows 5 skip-stop bays along the corridor, each bay containing 7-11 non-transfer stops for each route. Note that the last skip-stop bay is much longer than the other bays, and in that bay the numbers of non-transfer stops for route 1 and route 2 are not equal (there are 11 non-transfer stops for route 1 and 10 non-transfer stops for route 2). These small anomalies were created by the recipe for generating the real design; see step 2 in Section 2.2.3 for details.

The stop location plan of the optimal all-stop design for the same corridor under the same demand is shown in Figure 2.3b for comparison. The comparison between Figure 2.3a and b unveils high similarity: note that both $s(x)$ curves in the two figures peak at $x = 13$ and 28 km, and trough at $x = 0$ and 20 km. The two figures' only difference seems to be that the AB-type design exhibits smaller stop spacings in that the $s(x)$ curve of the AB-type design is like a scaled-down

version of that of the all-stop design. In addition, Figure 2.3c illustrates the optimal AB-type design for a higher demand density ($\Lambda_c/L = \Lambda_{cc}/L = 75$ trips/km/h) and the same demand variation. For this high demand case, the optimal design features three routes per direction, and the non-transfer stops of route 3 are marked by "+". Note that the stop spacing curve in Figure 2.3c is also similar to the one in Figure 2.3a, but is even 'flatter'. Similar findings are also observed when comparing the optimal AB-type and all-stop designs for the same bus corridor but with a less varied origin density ($\sigma_o = 8$ km); see Figure 2.3d and e.

The similarities between the optimal $s(x)$ curves in these figures can be explained by examining the following approximation of (2.32) under symmetric demand and design (i.e. $P_c(x) = P_{cc}(x)$, $Q_c(x) = Q_{cc}(x)$, $o_c(x) = o_{cc}(x)$, $r_c = r_{cc}$ and $H_c = H_{cc}$):

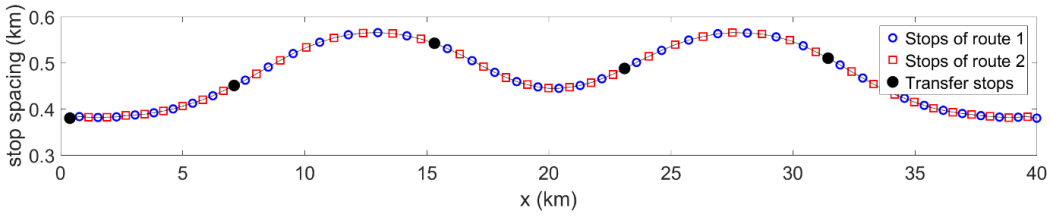
$$\tilde{s}(x) = \sqrt{\frac{\frac{k_c(x)+1}{r_c k_c(x)+1}\left(o_c(x)+\frac{\pi_m}{\mu H_c}\right)\tau+\frac{k_{cc}(x)+1}{r_{cc} k_{cc}(x)+1}\left(o_{cc}(x)+\frac{\pi_m}{\mu H_{cc}}\right)\tau+\frac{\pi_s}{\mu}}{\left(P_c(x)+Q_c(x)+P_{cc}(x)+Q_{cc}(x)\right)/4v_w+\left(r_c k_c(x)+1\right)\left(b_c(x)+b_{cc}(x)\right)/3v}} \approx$$

$$\sqrt{\frac{\left(\frac{k_c(x)+1}{r_c k_c(x)+1}o_c(x)+\frac{k_{cc}(x)+1}{r_{cc} k_{cc}(x)+1}o_{cc}(x)\right)\tau+\frac{\pi_s}{\mu}}{\left(P_c(x)+Q_c(x)+P_{cc}(x)+Q_{cc}(x)\right)/4v_w}} \approx \sqrt{\frac{\frac{2}{r_c}o_c(x)\tau}{\left(P_c(x)+Q_c(x)\right)/2v_w}} \tag{2.37}$$

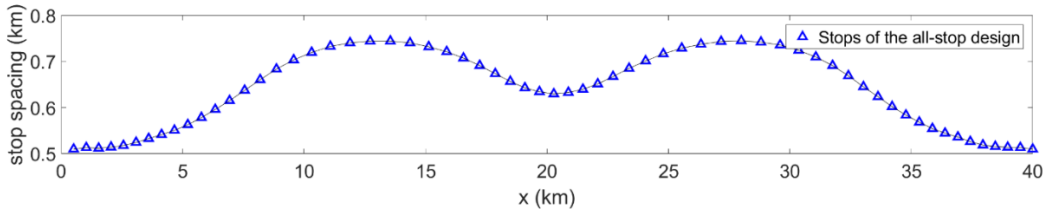
This approximation is obtained by ignoring the relatively minor terms under the square root sign, including: $\frac{\pi_m}{\mu H_c}$ (which is usually much smaller than $o_c(x)$[6]), $\frac{r_c-1}{r_c T(x)}$ (which is much smaller than $\frac{1}{r_c}$ in general), $\frac{\pi_s}{\mu}$ (which is small, especially for buses) in the numerator, and $2T(x)b_c(x)/3v$ in the denominator (since the backtracking trip density $b_c(x)$ is very small). By removing these minor terms, it can be clearly seen that optimal stop spacing is mainly determined by the ratio between on-board patron flow and the sum of origin and destination densities, i.e., $\frac{o_c(x)}{P_c(x)+Q_c(x)}$. In other words, the shape of $\sqrt{\frac{o_c(x)}{P_c(x)+Q_c(x)}}$ $(x \in [0, L])$ dictates the

[6] To see why, note that π_m mainly consists of the staff wage component. Hence, $\frac{\pi_m}{\mu}$ is a proxy of the number of staff assigned to each transit vehicle, and $\frac{\pi_m}{\mu H_c}$ can be interpreted as the "flow of staff" that passes by any point (e.g. x) in the corridor. This flow of staff is usually much smaller than the flow of on-board patrons passing by x, i.e., $o_c(x)$.

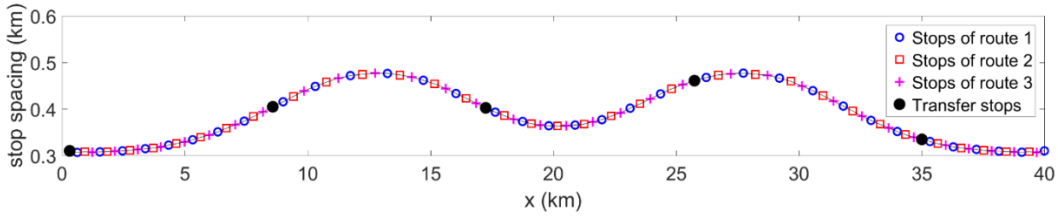
shape of the optimal $s(x)$. It is not very surprising to see that the above finding is consistent with Wirasinghe and Ghoneim's (1981) for optimal all-stop designs. Moreover, this approximation unveils that optimal stop spacing will scale down as r_c increases. This explains why the $s(x)$ curves for the all-stop and AB-type designs under the same demand pattern (although perhaps with different demand density levels) look very similar, and why the curve becomes flatter as the number of skip-stop routes increases. Note that a lesser degree of similarity will be observed between the all-stop and AB-type service designs of rail corridors due to the much larger π_s.



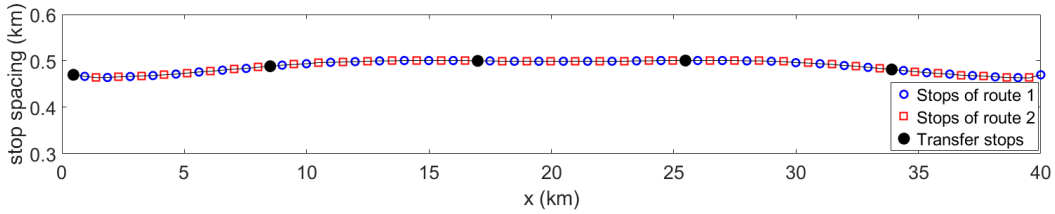
(a) AB-type design under demand of low density and high spatial variation ($\sigma_o = 4$ km, $\Lambda_c/L = \Lambda_{cc}/L = 37.5$ trips/km/h).



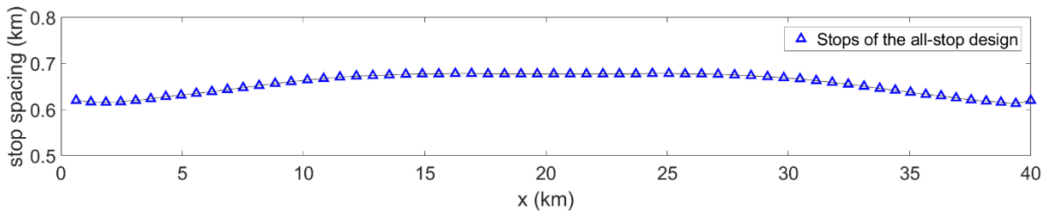
(b) All-stop design under demand of low density and high spatial variation ($\sigma_o = 4$ km, $\Lambda_c/L = \Lambda_{cc}/L = 37.5$ trips/km/h).



(c) AB-type design under demand of medium density and high spatial variation ($\sigma_o = 4$ km, $\Lambda_c/L = \Lambda_{cc}/L = 75$ trips/km/h).



(d) AB-type design under demand of low density and low spatial variation ($\sigma_o = 8$ km, $\Lambda_c/L = \Lambda_{cc}/L = 37.5$ trips/km/h).



(e) All-stop design under demand of low density and low spatial variation ($\sigma_o = 8$ km, $\Lambda_c/L = \Lambda_{cc}/L = 37.5$ trips/km/h).

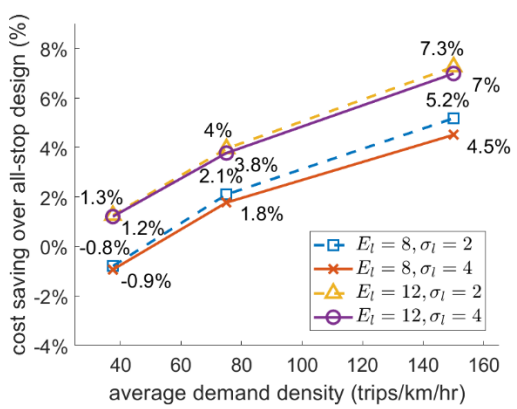
Figure 2.3 Optimal stop spacings for a bus corridor with symmetric demand.

2.3.4 Parametric analysis for symmetric AB-type designs

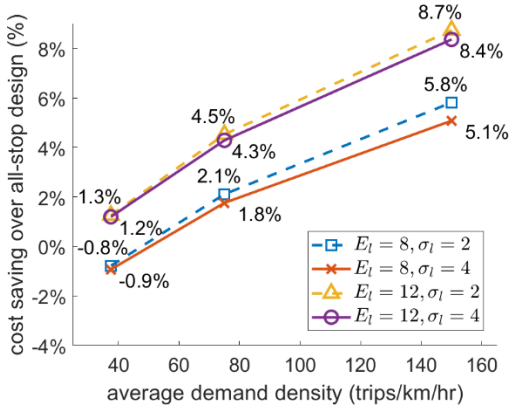
This section examines all the 144 symmetric-demand instances. To demonstrate the AB-type design's effectiveness, the generalized costs of the optimal AB-type design are compared for each instance *with two or more routes in each direction* and the optimal all-stop design. The generalized cost savings of the AB-type design are plotted against the key parameters in Figure 2.4a-d and Figure 2.5a-b.

Specifically, Figure 2.4a-c present the generalized cost savings of the AB-type design over the all-stop design for a bus corridor in a high-wage city with $\sigma_o = \infty$, 8, and 4 km, respectively. Each figure presents four percentage cost saving curves against the average demand density for the cases of $\sigma_l \in \{4, 8\}$ km and $E_l \in \{8, 12\}$ km, respectively. All these curves show that the cost saving increases with demand density. Negative cost savings (as low as -0.9%) are found in all the three figures when the demand density (37.5 trips/km/h) and the mean trip length (8 km) are both small, indicating that the all-stop design outperforms the AB-type design. On the other hand, cost savings of up to 8-9% are observed when the demand density is high (150 trips/km/h); see Figure 2.4b and c. Note in Figure 2.4c that the points representing the highest demand density (150 trips/km/h) and the largest mean trip length (12 km) are missing, because the vehicle capacity constraints (2.22d-e) are violated under these high-demand cases. Comparing the four curves in each figure reveals that: i) cost saving increases with the mean trip length; and ii) trip length variation has a marginal negative effect on cost saving. These findings are as expected: the AB-type design benefits patrons (especially those with longer trip distances) but raises agency side costs, hence the cost advantage of the AB-type design will grow as the system serves more patron-kms of travel.

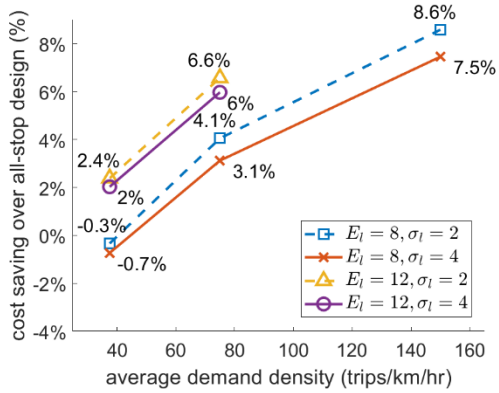
Comparison across the three figures shows that cost saving also increases with the spatial variation in trip origin density (which is an indicator of the demand heterogeneity). This means the AB-type service becomes more competitive when the demand is more heterogeneous.



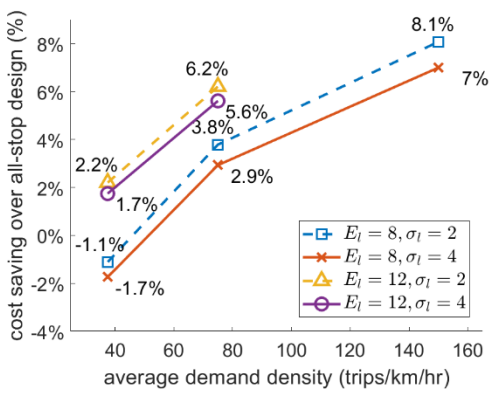
(a) Uniform demand in a high-wage city ($\sigma_o = \infty, \mu = 20$ \$/h).



(b) Less-spatially-varied demand in a high-wage city ($\sigma_o = 8$ km, $\mu = 20$ \$/h).



(c) Highly-spatially-varied demand in a high-wage city ($\sigma_o = 4$ km, $\mu = 20$ \$/h).



(d) Highly-spatially-varied demand in a low-wage city ($\sigma_o = 4$ km, $\mu = 5$ \$/h).

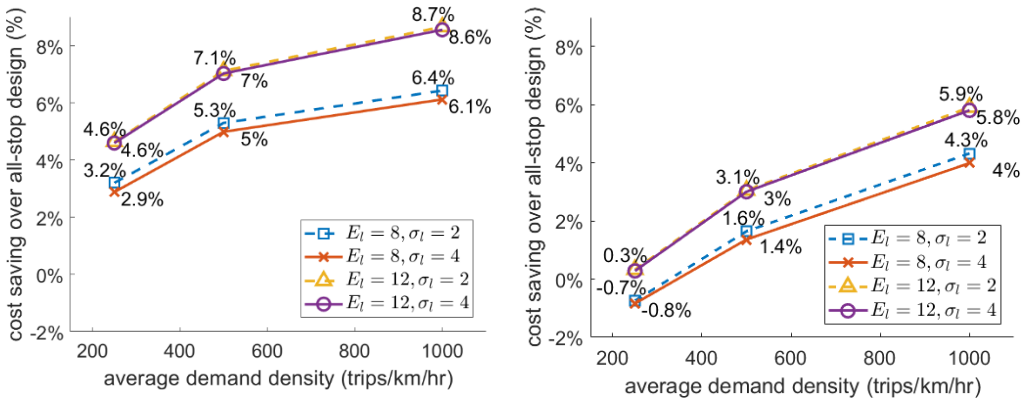
Figure 2.4. Percentage savings in AB-type bus systems' generalized cost.

To examine the effect of the value of time, we plot in Figure 2.4d the same four cost saving curves with the same parameter values as in Figure 2.4c, but for a low-wage city ($\mu = 5$ \$/h). Comparising Figure 2.4c and d reveals that the AB-type design is slightly more favorable in high-wage cities. This is also as expected, because in a high-wage city the patrons' travel times are considered more valuable. Similar results are also observed under other demand patterns, omitted here for brevity.

Similar results are observed for AB-type rail systems; see e.g. Figure 2.5a and b for the percentage cost saving curves with high spatially-varied demand

($\sigma_o = 4$ km) km in high- and low-wage cities, respectively. Of note, these two figures show that the AB-type rail system is significantly more advantageous in high-wage cities. In addition, comparing Figure 2.4c-d and Figure 2.5a-b reveals that the AB-type scheme is more beneficial when applied to a bus corridor than to a rail corridor. Note in Figure 2.4c that the cost saving reaches 8.6% for a bus corridor in a high-wage city, when the average demand density is only 150 trips/km/hr (with $E_l = 8$ km and $\sigma_l = 2$ km). On the other hand, Figure 2.5a shows for the same trip length distribution and value of time AB-type service rail corridor's cost saving is only 6.4% even at a very high demand density of 1000 trips/km/hr. This is because the AB-type design requires operation of more transit routes, and adding rail routes is much more expensive than adding bus routes.

Note that the above findings with regard to effects of mean trip length, demand heterogeneity, and the value of time on AB-type design performance are consistent with that reported in the literature (e.g. Larrain and Muñoz, 2016).



(a) A high-wage city ($\mu = 20$ \$/h) (b) A low-wage city ($\mu = 5$ \$/h).

Figure 2.5. Percentage savings in AB-type rail systems' generalized cost with highly-spatially-varied demand.

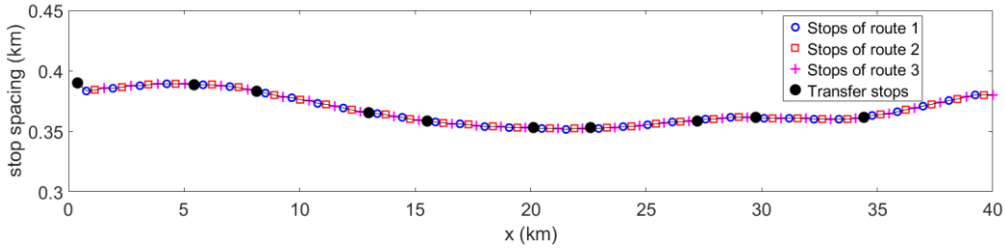
2.3.5 Designs under asymmetric demand

We first consider A bus corridor in a high-wage city ($\mu = 20$ \$/h) where the demand densities of the two travel directions are different; i.e. $\frac{\Lambda_c}{L} = 150$ trips/km/h and $\frac{\Lambda_{cc}}{L} = 75$ trips/km/h are considered here. The distributions of the trip origins and the trip lengths are assumed to be the same between the two directions with $\sigma_o = 8$ km, $E_l = 8$ km, and $\sigma_l = 4$ km. We find that the optimal

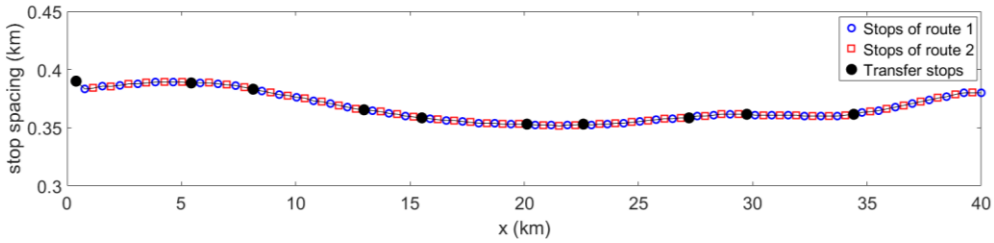
AB-type design is also asymmetric. Specifically, it features a 3-route AB-type service in the clockwise direction and a 2-route counterclockwise AB-type service. The optimal stop location plans for the two directions are presented in Figure 2.6a and b, respectively, where the transfer stops are marked by solid black circles, and the non-transfer stops of different routes in each direction are marked by blue circles, red squares, and red crosses, respectively. Note that the two directions share the same set of stops, and the same set of transfer stops. However, non-transfer stops are allocated to three routes in the clockwise direction, and to only two counterclockwise routes.

For comparison, the AB-type design for the same corridor with lower but still asymmetric demand densities ($\frac{\Lambda_c}{L} = 75$ trips/km/h and $\frac{\Lambda_{cc}}{L} = 37.5$ trips/km/h) in a high-wage city was also optimized. This new optimal design features a 2-route clockwise AB-type service (see Figure 2.6c) and an all-stop counterclockwise service. This all-stop service plan is not shown because it uses the same set of stops as the clockwise direction. We further show the optimal design for the same corridor with the lower demand densities in a low-wage city ($\mu = 5$ \$/h) in Figure 2.6d. Note in this last case that the all-stop service becomes optimal in both directions, despite the asymmetric demand.

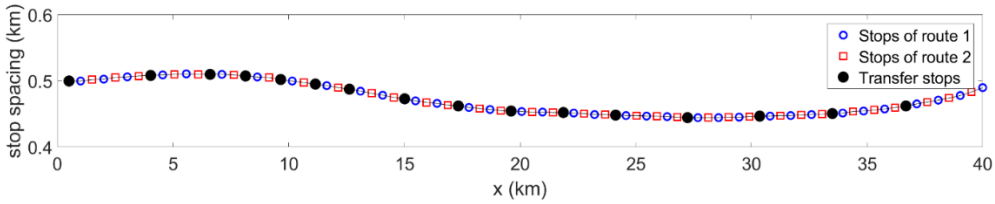
Comparing Figure 2.6a-d confirms that the optimal number of skip-stop routes still increases with the demand level and the value of time, and that more routes are often required in the higher demand direction. However, the similarity between the $s(x)$ curves in these figures are not clear.



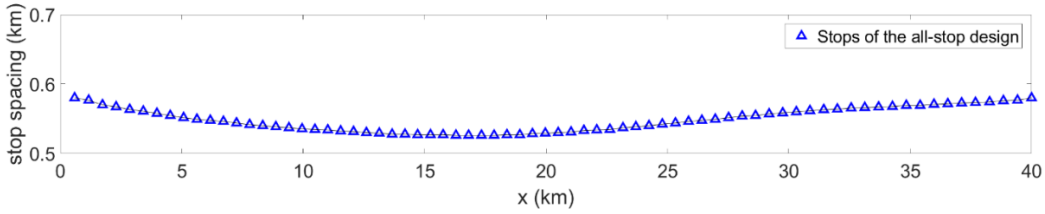
(a) Clockwise AB-type design in a high-wage city with a higher demand level ($\frac{\Lambda_c}{L} = 150$ trips/km/h, $\frac{\Lambda_{cc}}{L} = 75$ trips/km/h, $\mu = 20$ \$/h)



(b) Counterclockwise AB-type design in a high-wage city with a higher demand level ($\frac{\Lambda_c}{L} = 150$ trips/km/h, $\frac{\Lambda_{cc}}{L} = 75$ trips/km/h, $\mu = 20$ \$/h)



(c) Clockwise AB-type design in a high-wage city with a lower demand level ($\frac{\Lambda_c}{L} = 75$ trips/km/h, $\frac{\Lambda_{cc}}{L} = 37.5$ trips/km/h, $\mu = 20$ \$/h).



(d) All-stop design in both directions in a low-wage city with a lower demand level ($\frac{\Lambda_c}{L} = 75$ trips/km/h, $\frac{\Lambda_{cc}}{L} = 37.5$ trips/km/h, $\mu = 5$ \$/h).

Figure 2.6 Optimal stop spacings for a bus corridor under asymmetric demands.

2.4 Summary of the AB-Type Design Model

This chapter presents CA models formulated for optimizing the AB-type service design (including both the stop locations and the routing plan) in loop transit corridors under arbitrary heterogeneous demand. This design allows the stop spacing and the number of non-transfer stops between two consecutive transfer stops to vary over the corridor, and the numbers of skip-stop routes to be different between the two service directions. This work is an important extension of the previous work (Gu et al., 2016), which assumed uniformly distributed demand.

The CA optimization problem was solved by partially decomposing the formulation by the local spatial coordinate (i.e., $x \in [0, L)$), and optimizing the decision functions for each x. The undecomposable parts pertaining to

backtracking trips were taken as given in the partial decomposition, and updated through an iterative process. Numerical tests demonstrate that our solution approach produced near-optimal solutions in very short times.

By addressing the demand heterogeneity, this work also paves the way for real world application of CA models in AB-type transit corridor design. These models' practicality is further enhanced by our proposed recipe for transforming the CA solutions to realistic transit service designs.

The optimal AB-type design's advantage over the conventional all-stop design was manifested in the large array of numerical experiments. Generalized cost savings of up to 8-9% were observed. Note that the real cost savings for bus corridors can be even larger, if the arrival times of some buses of different routes operating in the same direction are coordinated at transfer stops (please refer to Gu et al., 2016, for more details on such coordination strategy).

Chapter 3 Local-Express Designs

In this chapter, a CA model is developed for optimizing the local-express service in a linear corridor. The model imbeds patron route choice behaviors through a user-optimal route assignment process. In addition to a single-mode system, such as a bus or rail system, this proposed model can also be applied to a bimodal system, consisting, for example, of both a local bus line and an express rail line. The model can jointly optimize stop locations and the routing plan under arbitrary heterogeneous demand patterns. By integrating the calculus of variations method into a three-stage iterative process, we show that this model can be efficiently solved to near-optimality. Numerical analysis also shows that the optimal local-express design outperforms conventional all-stop design under various operating conditions. Specifically, a bimodal local-express system is shown to be more efficient than a single-mode system.

Section 3.1 presents the formulations of the local-express design. Section 3.2 describes the solution method of the proposed models. Section 3.3 provides numerical analysis. Section 3.4 summarizes local-express design models.

3.1 The CA Formulation

Consider a linear corridor of length L (km). The demand is assumed to be exogenous, and its density is represented by a slow-varying, integrable function $\lambda(x, y)$ (trips/km^2/h), where x and y are origin and destination location coordinates of a trip, respectively; $x, y \in [0, L]$.

For a corridor's two directions, supposing the eastbound direction is from coordinate 0 to L and the westbound direction is the opposite, as indicated by subscript $d \in \{E, W\}$), the trip origin and destination densities ($P^d(x), Q^d(x)$, trips/h/km) at any location x of the corridor, are expressed as follows:

(1) Eastbound ($d = E$) trip origin density at location x is $P^E(x) = \int_{y=x}^{L} \lambda(x, y)dy$, which implies the sum of OD pairs originating from location x to any locations to the east of x in the corridor. Trip destination density at location x

is $Q^E(x) = \int_{y=0}^{x} \lambda(y,x)dy$, which is the sum of OD pairs originating from the west to location x.

(2) Similarly, westbound ($d = W$) trip origin and destination densities at location x are $P^W(x) = \int_{y=0}^{x} \lambda(x,y)dy$, $Q^W(x) = \int_{y=x}^{L} \lambda(y,x)dy$, respectively.

The local-express model involves two scalar decision variables, the headways H_l and H_e for the local and express transit, respectively; and two decision functions, i.e., the stop density functions $\delta_l(x)$ and $\delta_e(x)$ of location x for the local and express transit, respectively. We assume that local and express stops can be located at any point of the corridor; see in Figure 1.2. $\delta_l(x), \delta_e(x)$ are continuous with slow changes in space. The headways are identical between the two directions during the study period (e.g., peak hours). We assume patrons always walk to the nearest stop to access the transit service and their route choice behaviors follow a deterministic manner: they always choose the route with the minimum travel time. We also assume that patrons always travel forward so that there are no backtracking trips in the system.

We first define five route types in the corridor and present the patron's route choice model in Section 3.1.1. Then Sections 3.1.2 and 3.1.3 summarize patron and agency cost components, respectively. Section 3.1.4 presents the problem formulation. For simplicity, the formulations in Section 3.1.1 are derived for the eastbound direction only. Similar models can easily be formulated for the other direction.

3.1.1 Route assignment

There are four groups of OD pairs according to the type of the origin and destination stops, i.e. whether they are local or express stops (see Figure 1.2). Group 1 includes trips with both origin and destination falling in the neighborhood of express stops; group 2 includes trips originating from local stops and ending at express stops; group 3 includes trips originating from express stops and ending at local stops; and group 4 includes trips both originating from and ending at local stops.

Different route options are associated with the four OD pairs groups: apparently, patrons in Group 1 will choose the express line only; this route option

is denoted by e. Patrons in group 2 have two route options: i) they may choose local transit only (denoted by l) to complete their trips without a transfer; or ii) they could take the local line first and then transfer to the express line to reach their destination (denoted by le). Similarly, patrons in group 3 also have two route options: l and el; the latter representing a route that takes the express line first and then transfers to the local line. Lastly, patrons in group 4 also have two route options: l and lel; the latter representing a route that takes the local line first, then transfers to the express line at the nearest express stop, and finally transfers back to the local line to complete the trip. In total, there are 5 route types of s, denoted by the set $\mathcal{R} = \{e, l, le, el, lel\}$.

According to our assumption, patrons will always choose to take the fastest route. Note that the local-only route l is available for all patrons. Under this assumption, route choice can be derived by comparing the travel time of another route option (el, le, lel) against that of l.

Travel time is then derived for each route. A trip's travel time consists of four parts: i) access and egress time by walking to/from the nearest transit stop; ii) the total waiting time at the origin stop and transfer stop(s) (if any); iii) the total in-vehicle travel time; and iv) the transfer penalty if the trip involves transfers. Detailed formulas for the travel times of the 5 route types are given below:

$$t^E_{(e)}(x,y) = \kappa(x,y) + \frac{H_e}{2} + \int_x^y \frac{1}{V_e^E(u)}du \tag{3.1}$$

$$t^E_{(l)}(x,y) = \kappa(x,y) + \frac{H_l}{2} + \int_x^y \frac{1}{V_l^E(u)}du \tag{3.2}$$

$$t^E_{(le)}(x,y) = \kappa(x,y) + \frac{H_l+H_e}{2} + \left(\int_x^y \frac{1}{V_e^E(u)}du + \gamma^E(x)\right) + C_t \tag{3.3}$$

$$t^E_{(el)}(x,y) = \kappa(x,y) + \frac{H_l+H_e}{2} + \left(\int_x^y \frac{1}{V_e^E(u)}du + \gamma^E(y)\right) + C_t \tag{3.4}$$

$$t^E_{(lel)}(x,y) = \kappa(x,y) + \frac{2H_l+H_e}{2} + \left(\int_x^y \frac{1}{V_e^E(u)}du + \gamma^E(x) + \gamma^E(y)\right) + 2C_t \tag{3.5}$$

where $\kappa(x,y) = \frac{1}{4v_w}\left(\frac{1}{\delta_l(x)} + \frac{1}{\delta_l(y)}\right)$ is the (average) access and egress time; v_w is the patrons' average walking speed (km/h). The second terms in the right-hand-side (RHS) of (3.1-3.5) are the average total waiting time at the origin and transfer

stops. The third terms in the RHS of (3.1-3.5) are the in-vehicle travel time, where $V_e^E(u)$ and $V_l^E(u)$ are the commercial speeds at location u of the express and local transit vehicles, respectively, which are given by:

$$\frac{1}{V_e^E(u)} = \frac{1}{v_e} + \tau_e^0 \delta_e(u) + \max(B_e^E(u)\tau_e^b, A_e^E(u)\tau_e^a)\, H_e \tag{3.6}$$

$$\frac{1}{V_l^E(u)} = \frac{1}{v_l} + \tau_l^0 \delta_l(u) + \max\left(B_l^E(u)\tau_l^b, A_l^E(u)\tau_l^a\right) H_l \tag{3.7}$$

where for express transit vehicles, v_e is the cruise speed; τ_e^0 is the fixed delay per stop; τ_e^b and τ_e^a are the boarding and alighting delays per patron, respectively. Note that $B_e^E(u)$ and $A_e^E(u)$ are the boarding and alighting densities (including boardings and alightings occurring when the patrons make transfers) for the express line, respectively; and $B_l^E(u)$ and $A_l^E(u)$ are the boarding and alighting densities for the local line, respectively. These will be introduced in equation (3.11).

The $f^E(x)$ in (3.3) and (3.5) and $\gamma^E(y)$ in (3.4) and (3.5) represent the additional travel times experienced on the local line. Here the local-line travel distance on the origin side of a specific trip of route le is approximated by its mean, i.e., $\frac{1}{2\delta_e(x)}$. Hence, the additional travel time on the local line for route le can be expressed as $\gamma^E(x) = \frac{1}{2\delta_e(x)}\left(\frac{1}{V_l^E(x)} - \frac{1}{V_e^E(x)}\right)$. The same additional travel time should be added on the destination side of route el, and on both origin and destination sides of route lel.

The last terms in the RHS of (3.3-3.5) are the transfer penalties between local and express lines, where C_t denotes the penalty per transfer (h/transfer).

Therefore, for each of OD groups 2, 3, and 4, there exists a critical travel distance $C_{(I)}^E(x), \forall I \in \{le, el, lel\}$, in which a patron in a corresponding OD group will choose route I if the trip distance is greater than $C_{(I)}^E(x)$. Specifically, the following binary variable is defined:

$$\Omega_{(I)}^E(x,y) = \begin{cases} 1 & y - x > C_{(I)}^E(x) \\ 0 & y - x \le C_{(I)}^E(x) \end{cases}, \forall I \in \{le, el, lel\} \tag{3.8}$$

where $\Omega_{(I)}^{E}(x,y)$ takes 1 if route I is selected. For notation completeness, $\Omega_{(e)}^{E}(x,y)=1$ is specified for all x and y. The critical distances $C_{(I)}^{E}(x)$ can be solved from the following equations:

$$t_{(l)}^{E}\left(x,x+C_{(I)}^{E}(x)\right)=t_{(I)}^{E}\left(x,x+C_{(I)}^{E}(x)\right),\forall I\in\{le,el,lel\} \tag{3.9}$$

where $t_{(l)}^{E}$ and $t_{(I)}^{E},\forall I\in\{le,el,lel\}$ are given by equations (3.2-3.5). $C_{(I)}^{E}(x)$ can be easily solved when the design functions/variables $(\delta_e(x),\delta_l(x),H_e,H_l)$ are given.

Each route's demand, i.e. $\lambda_{(I)}(x,y),\forall I\in\mathcal{R}$, is then calculated by the following equations:

$$\lambda_{(e)}^{E}(x,y)=\lambda(x,y)p_e(x)p_e(y)\Omega_{(e)}^{E}(x,y) \tag{3.10a}$$

$$\lambda_{(le)}^{E}(x,y)=\lambda(x,y)p_l(x)p_e(y)\Omega_{(le)}^{E}(x,y) \tag{3.10b}$$

$$\lambda_{(el)}^{E}(x,y)=\lambda(x,y)p_e(x)p_l(y)\Omega_{(el)}^{E}(x,y) \tag{3.10c}$$

$$\lambda_{(lel)}^{E}(x,y)=\lambda(x,y)p_l(x)p_l(y)\Omega_{(lel)}^{E}(x,y) \tag{3.10d}$$

$$\lambda_{(l)}^{E}(x,y)=\begin{cases}\lambda(x,y)-\sum_{I\in\{e,le,el,lel\}}\lambda_{(I)}^{E}(x,y), & x\le y\\ 0, & x>y\end{cases} \tag{3.10e}$$

where $p_e(u)$ denotes the probability that location u is closer to an express stop than a local one, i.e. $p_e(u)=\frac{\delta_e(u)}{\delta_l(u)}$. Also, $p_l(u)=1-p_e(u)$ denotes the probability that u is closer to a local stop.

Based on the route demand, $\lambda_{(I)}^{E}(x,y)$, the trip origin and destination densities for each route can be obtained, $P_{(I)}^{E}(u)=\int_{y=u}^{L}\lambda_{(I)}^{E}(u,y)dy$ and $Q_{(I)}^{E}(u)=\int_{x=0}^{u}\lambda_{(I)}^{E}(x,u)dx$. Then, the boarding and alighting densities (including transfer boarding) for local and express lines, $B_l^E(x),A_l^E(x)$ and $B_e^E(x),A_e^E(x)$, can be given by:

$$B_l^E(x)=P_{(l)}^{E}(x)+P_{(le)}^{E}(x)+P_{(lel)}^{E}(x)+\lambda_{e\to l}^{E}(x) \tag{3.11a}$$

$$A_l^E(x)=Q_{(l)}^{E}(x)+Q_{(el)}^{E}(x)+Q_{(lel)}^{E}(x)+\lambda_{l\to e}^{E}(x) \tag{3.11b}$$

$$B_e^E(x)=P_{(e)}^{E}(x)+P_{(el)}^{E}(x)+\lambda_{l\to e}^{E}(x) \tag{3.11a}$$

$$A_e^E(x) = Q_{(e)}^E(x) + Q_{(le)}^E(x) + \lambda_{e\to l}^E(x) \quad (3.11d)$$

where $\lambda_{e\to l}^E(x)$ and $\lambda_{l\to e}^E(x)$ are the transfer densities at x from express to local and from local to express lines, respectively. These are given by $\lambda_{e\to l}^E(x) = Q_{(el)}^E(x) + Q_{(lel)}^E(x)$ and $\lambda_{l\to e}^E(x) = P_{(le)}^E(x) + P_{(lel)}^E(x)$.

The above demand assignment process (3.8-3.11) will be iteratively updated until equilibrium is reached among all routes.

3.1.2 Users cost metrics

User metrics are measured in terms of travel time consisting of four components: (i) walking access and egress cost (UC_a), (ii) waiting cost at transit stops (UC_w), (iii) in-vehicle travel cost (UC_i), and (iv) transfer penalty (UC_t).

The total access and egress cost is the total walking time in the corridor, given by:

$$UC_a = \int_{x=0}^{L} \frac{1}{4v_w\delta_l(x)} \left(P^E(x) + Q^E(x) + P^W(x) + Q^W(x)\right) dx \quad (3.12)$$

where $P^E(x) + Q^E(x) + P^W(x) + Q^W(x)$ is the total trip origin and destination density at location x; v_w is patrons' average walking speed; and $\frac{1}{4\delta_l(x)}$ is the average walking distance to/from the closest stop at x.

The total waiting cost, UC_w, is the sum of the waiting times at the trips' origin and transfer stops. This can be expressed as:

$$UC_w = \int_{x=0}^{L} \left(B_l^E(x) + B_l^W(x)\right) \frac{H_l}{2} dx + \int_{x=0}^{L} \left(B_e^E(x) + B_e^W(x)\right) \frac{H_e}{2} dx \quad (3.13)$$

where the two integrals are the collective waiting times at local and express lines, respectively; and $\frac{H_l}{2}$ and $\frac{H_e}{2}$ are the average waiting times for a local line vehicle and an express vehicle, respectively. Here it is assumed that: i) patrons' arrivals at transit stops are uniformly distributed over the time; ii) headways between transit vehicles are deterministic (i.e. the stochastic variations are ignored); and iii) the two transit modes' operations are not coordinated at the express stops.

The in-vehicle travel cost, UC_i, is the sum of in-vehicle travel costs in express and local lines. This can be estimated by:

$$UC_i = \int_{x=0}^{L} \sum_{d\in\{E,W\}} \left(\frac{o_e^d(x)}{V_e^d(x)} + \frac{o_l^d(x)}{V_l^d(x)} \right) dx \tag{3.14}$$

where $o_e^d(x)$ and $o_l^d(x)$ (trips/h), $d \in \{E, W\}$ are the on-board patron flows on the express and local lines, respectively, which are given by:

$$o_l^d(x) = \int_{u=0}^{x} \int_{y=x}^{L} \lambda_{(l)}^d(u,y) du dy + \frac{\lambda_t^d(x)}{2\delta_e(x)} \tag{3.15}$$

$$o_e^d(x) = o^d(x) - o_l^d(x) \tag{3.16}$$

The second term in the RHS of (3.15) accounts for the transfer patron flow at x, in which $\lambda_t^d(x) \equiv \lambda_{e\to l}^d(x) + \lambda_{l\to e}^d(x)$ denotes the total transfer demand density. In (3.16), $o^d(x)$ is the overall on-board patron flow at x for direction d, given by $o^E(x) = \int_{u=0}^{x} \int_{y=x}^{L} \lambda(u,y) du dy$ and $o^W(x) = \int_{u=x}^{L} \int_{y=0}^{x} \lambda(u,y) du dy$.

Finally, the transfer penalty, UC_t, depends on the number of transfers of the five route types s, and is given by:

$$UC_t = \int_{x=0}^{L} \left(\lambda_t^E(x) + \lambda_t^W(x) \right) C_t dx \tag{3.17}$$

3.1.3 Agency cost metrics

The agency cost consists of: i) distance-based vehicle operating cost (for example fuel cost), AC_K; ii) time-based vehicle operating cost (such as amortized vehicle purchase cost and staff wages), AC_H; iii) amortized line infrastructure cost (for example busway or rail tracks), AC_I; and iv) the amortized stop infrastructure cost, AC_S. These are formulated as:

$$AC_K = \frac{\pi_v^l}{\mu} \frac{2L}{H_l} + \frac{\pi_v^e}{\mu} \frac{2L}{H_e} \tag{3.18}$$

$$AC_H = \frac{\pi_m^l}{\mu H_l} \int_0^L \left(\frac{1}{V_l^E(x)} + \frac{1}{V_l^W(x)} \right) dx + \frac{\pi_m^e}{\mu H_e} \int_0^L \left(\frac{1}{V_e^E(x)} + \frac{1}{V_e^W(x)} \right) dx \tag{3.19}$$

$$AC_I = \frac{2\pi_i^l L}{\mu} + \frac{2\pi_i^e L}{\mu} \tag{3.20}$$

$$AC_S = \frac{\pi_s^l}{\mu} \int_0^L \delta_l(x) dx + \frac{\pi_s^e}{\mu} \int_0^L \delta_e(x) dx \tag{3.21}$$

where μ is the value of time (\$/h); π_v^l and π_m^l are the unit operating costs per vehicle-km and per vehicle-hour of local transit service for each service hour, respectively; π_i^l and π_s^l are the unit infrastructure construction and maintenance

costs per km of local line and per local stop, respectively, amortized for each service hour of local line; and similarly, $\pi_v^e, \pi_m^e, \pi_i^e$ and π_s^e denote the express service's operating and infrastructure cost coefficients .

3.1.4 Problem formulation

The generalized cost optimization problem is then formulated as:

$$\min SC = UC_a + UC_w + UC_i + UC_i + AC_K + AC_H + AC_I + AC_S \quad (3.22a)$$

subject to:

$$H_l \geq H_l^{min} \quad (3.22b)$$

$$H_e \geq H_e^{min} \quad (3.22c)$$

$$\frac{K_l}{H_l} \geq \max_{x}\{o_l^E(x), o_l^W(x)\} \quad (3.22d)$$

$$\frac{K_e}{H_e} \geq \max_{x}\{o_e^E(x), o_e^W(x)\} \quad (3.22e)$$

$$\delta_e(x) \leq \delta_l(x), \forall x \in (0, L] \quad (3.22f)$$

$$\delta_e(x), \delta_l(x) \geq 0, \forall x \in (0, L] \quad (3.22g)$$

where the cost components in the objective function (3.22a) are given by equations (3.12-3.21); H_l^{min} and H_e^{min} are the local and express lines' minimum service headways, respectively; K_l and K_e are the local and express vehicles' passenger-carrying capacity , respectively. Constraints (3.22d-e) ensure the local and express vehicles' onboard patron numbers never exceed their corresponding vehicle capacities. Constraints (3.22f) indicate local stops are more densely deployed than express stops.

3.2 Solution Method

Section 3.2.1 presents the solution approach to the local-express service model. Section 3.2.2 furnishes a recipe for converting the optimal CA model solution to a real local-express design.

3.2.1 Solution to the local-express model

The optimization problem (3.22) has two scalar decision variables H_l and H_e and two decision functions $\delta_e(x)$ and $\delta_l(x)$. The problem can be solved via calculus of variations since the part of the objective function (3.22a) related to $\delta_e(x)$ and $\delta_l(x)$ is separable by x.

By inspecting the cost components (3.12-3.21), the objective function (3.22a) can be expressed as the sum of two parts: the first part being only related to the scalar decision variables H_l and H_e, and the second part being an integral with respect to x over $(0, L]$, whose integrand is related to $\delta_e(x)$ and $\delta_l(x)$. The detail is presented as follows:

$$SC = T(H_l, H_e) + \int_{x=0}^{L} G(H_l, H_e, \delta_e(x), \delta_l(x), x)dx \tag{3.23}$$

where,

$$T(H_l, H_e) \equiv \frac{\pi_v^l}{\mu}\frac{2L}{H_l} + \frac{\pi_v^e}{\mu}\frac{2L}{H_e} + \frac{2\pi_i^l L}{\mu} + \frac{2\pi_i^e L}{\mu} \tag{3.24}$$

$$G(H_l, H_e, \delta_e(x), \delta_l(x), x) \equiv \sum_{d\in\{E,W\}}\left[\frac{P^d(x)+Q^d(x)}{4v_w\delta_l(x)} + \frac{B_l^d(x)H_l}{2} + \frac{B_e^d(x)H_e}{2} + \frac{o_e^d(x)}{V_e^d(x)} + \frac{o_l^d(x)}{V_l^d(x)} + \lambda_t^d(x)C_t + \frac{\pi_m^l}{\mu H_l}\frac{1}{V_l^d(x)} + \frac{\pi_m^e}{\mu H_e}\frac{1}{V_e^d(x)} + \frac{\pi_s^l}{\mu}\delta_l(x) + \frac{\pi_s^e}{\mu}\delta_e(x)\right] \tag{3.25}$$

Note in (3.25) that values of $B_l^d(x), B_e^d(x), o_e^d(x), o_l^d(x), V_e^d(x), V_l^d(x)$ and $\lambda_t^d(x)$ depend on the demand assignment results, which are related to the decision variables/functions $H_l, H_e, \delta_e(x)$ and $\delta_l(x)$. Hence, a three-stage iterative algorithm is proposed next to solve the problem.

In the first stage, the demand assignment is conducted when the design is fixed by the given values of $H_l, H_e, \delta_e(x)$ and $\delta_l(x)$. The demand assignment is also processed via an iterative method. Assuming an initial split of demand for the five routes, represented as $\left\{\lambda_{(I)}^d(x,y)^{(0)}\middle| I \in \mathcal{R}, d \in \{E,W\}\right\}$ from origin x to destination y, then it is easy to calculate $B_l^d(x), B_e^d(x), V_e^d(x), V_l^d(x), \forall d \in \{E,W\}$ from (3.11a-d) and (3.6-3.7), and $t_{(I)}^d, \Omega_{(I)}^d(x,y), \forall I \in \mathcal{R}, d \in \{E,W\}$ from (3.1-3.5) and (3.8-3.9). Hence, the new demand split $\bar{\lambda}_{(I)}^d(x,y)$ can be obtained by (3.10a-e). After that, the demand split is updated using the method of successive averages (MSA) (Sheffi, 1985), specifically:

$$\lambda_{(l)}^{d}(x,y)^{(n)} = \lambda_{(l)}^{d}(x,y)^{(n-1)} + \frac{\bar{\lambda}_{(l)}^{d}(x,y) - \lambda_{(l)}^{d}(x,y)^{(n-1)}}{n} \tag{3.26}$$

where n is the time of the present iteration. This process is repeated until the assignment result converges.

In the second stage, H_l and H_e are fixed and the first stage's demand assignment results are used to optimize SC with respect to $\delta_e(x)$ and $\delta_l(x)$. Then the first-order conditions of SC with respect to $\delta_e(x), \delta_l(x)$ are used to obtain the following optimal $\delta_e^*(x), \delta_l^*(x)$:

$$\delta_e^*(x) = \sqrt{\frac{\sum_{d\in\{E,w\}} \frac{\lambda_t^d(x)}{2}\left(\frac{1}{V_l^d} - \frac{1}{v_e} - \max\left(B_e^d(u)\tau_e^b, A_e^d(u)\tau_e^a\right)H_e\right)}{\frac{\pi_s^e}{\mu} + \tau_e^0\left[\frac{2\pi_m^e}{\mu H_e} + o^E(x) + o^W(\mathrm{x}) - \int_{u=0}^{x}\left(\int_{y=x}^{L}\lambda_l^E(u,y)dy + \int_{y=x}^{L}\lambda_l^W(y,u)dy\right)du\right]}} \tag{3.27}$$

$$\delta_l^*(x) = \frac{1}{2\sqrt{v_w}}\sqrt{\frac{P^E(x) + Q^E(x) + P^W(x) + Q^W(x)}{\frac{\pi_s^l}{\mu} + \tau_l^0\left(\frac{2\pi_m^l}{\mu H_l} + o_l^E(x) + o_l^W(x)\right)}} \tag{3.28}$$

In the last stage, H_l and H_e are optimized based on the previous two stages' results. The first order conditions of SC with respect to H_l and H_e are inspected, which generate the unconstrained optimal $\widetilde{H}_e, \widetilde{H}_l$ as below.

$$\widetilde{H}_e = \sqrt{\frac{\frac{2\pi_v^e L}{\mu} + \frac{2\pi_m^e}{\mu}\int_0^L\left(\frac{1}{v_e} + \tau_e^0\delta_e(x)\right)dx}{\int_{x=0}^{L}\sum_{d\in\{E,W\}}\left(\frac{B_e^d(x)}{2} + o_e^d(x)\max\left(B_e^d(u)\tau_e^b, A_e^d(u)\tau_e^a\right)\right)dx}} \tag{3.29}$$

$$\widetilde{H}_l = \sqrt{\frac{\frac{2\pi_v^l L}{\mu} + \frac{2\pi_m^l}{\mu}\int_0^L\left(\frac{1}{v_l} + \tau_l^0\delta_l(x)\right)dx}{\int_{x=0}^{L}\sum_{d\in\{E,W\}}\left(\frac{B_l^d(x)}{2} + o_l^d(x)\max\left(B_l^d(u)\tau_l^b, A_l^d(u)\tau_l^a\right)\right)dx}} \tag{3.30}$$

Combing constraints (3.22b-e), the optimal H_e^*, H_l^* are finally given by:

$$H_e^* = \mathrm{mid}\left(H_e^{min}, \widetilde{H}_e, \frac{K}{\max\limits_{\mathrm{d}\in\{\mathrm{E,w}\},\mathrm{x}\in[0,\mathrm{L}]}\{o_e^d(x)\}}\right) \tag{3.31}$$

$$H_l^* = \mathrm{mid}\left(H_l^{min}, \widetilde{H}_l, \frac{K}{\max\limits_{\mathrm{d}\in\{\mathrm{E,w}\},\mathrm{x}\in[0,\mathrm{L}]}\{o_l^d(x)\}}\right) \tag{3.32}$$

where function $\mathrm{mid}(\cdot)$ returns the middle value of the three arguments. Note that the values of $B_l^d(x), B_e^d(x), o_e^d(x), o_l^d(x), V_e^d(x), V_l^d(x)$ and $\lambda_t^d(x)$ should be

updated between the stages. This three-stage algorithm runs by iteration until convergence is reached. The algorithm is summarized as follows:

Step 1. Discretize the continuous range of x, $(0, L]$, into m equal-sized intervals, each of length $\Delta x \equiv \frac{L}{m}$. Denote set $\mathcal{C}$ as $\mathcal{C} \equiv \left\{x_j \equiv \left(j - \frac{1}{2}\right)\Delta x \middle| j = 1,2, \dots, m\right\}$.

Step 2: Set feasible initial values for $H_l, H_e, \delta_e(x)$ and $\delta_l(x)$. Initialize the demand assignment by $\lambda^E_{(I)}(x,y)^{(0)} = \begin{cases} \frac{\lambda(x,y)}{5}, x < y \\ 0 \quad , x \geq y \end{cases}, \lambda^W_{(I)}(x,y)^{(0)} = \begin{cases} \frac{\lambda(x,y)}{5}, x > y \\ 0 \quad , x \leq y \end{cases}, \forall I \in \mathcal{R}$.

Step 3: Conduct demand assignment using (3.26) for each origin and destination pair $(x, y) \in \mathcal{C} \times \mathcal{C}$. Repeat the demand assignment process until $\left|\frac{\lambda^d_{(I)}(x,y)^{(n)} - \lambda^d_{(I)}(x,y)^{(n-1)}}{\lambda^d_{(I)}(x,y)^{(n-1)}}\right| \leq \varepsilon$ is satisfied for each (x, y), where ε is a pre-specified tolerance (e.g., $\varepsilon = 0.001$).

Step 4: Update the values of $B^d_l(x), B^d_e(x), o^d_e(x), o^d_l(x), V^d_e(x), V^d_l(x)$ and $\lambda^d_t(x)$ for each $x \in \mathcal{C}$, Find the optimal $\delta^*_e(x)$ and $\delta^*_l(x)$ using (3.27-3.28), and then update $B^d_l(x), B^d_e(x), o^d_e(x), o^d_l(x), V^d_e(x), V^d_l(x)$ and $\lambda^d_t(x)$ again.

Step 5: Find the optimal H^*_l and H^*_e using (3.31-3.32).

Step 6: Repeat steps (3-5) until $\left|\frac{H_e - H^*_e}{H_e}\right| + \left|\frac{H_l - H^*_l}{H_l}\right| \leq \varepsilon$ and $\left|\frac{\delta_e(x) - \delta^*_e(x)}{\delta_e(x)}\right| + \left|\frac{\delta_l(x) - \delta^*_l(x)}{\delta_l(x)}\right| \leq \varepsilon$ for each $x \in \mathcal{C}$.

3.2.2 Generating the exact stop locations

The solution obtained using the above algorithm consists of arrays of discrete point values on the continuous functions $\delta_e(x)$ and $\delta_l(x)$. A two-step recipe is proposed below to generate the exact locations for local stops and express stops. In Step 1, local stop locations are generated from the optimal $\delta_l(x)$ using a method similar to that proposed by Wirasinghe and Ghoneim (1981). In Step 2, the express stops are selected from the local stop set generated in Step 1 using the optimal $\delta_e(x)$.

Step 1. Place one stop at every x where $\int_{u=0}^{x} \frac{du}{\delta_l(u)} - \frac{1}{2}$ is an integer. Denote the resulting stop location plan as $\mathbf{\Omega} = \{x_i^S : i = 1, 2, \dots, N^S\}$, where N^S denotes the number of stops, and x_i^S the location of the i-th stop satisfying $0 = x_1^S < x_2^S < \cdots < x_{N^S}^S < L$.

Step 2. Mark $x_i^S \in \mathbf{\Omega}$ as an express stop if $\int_{u=0}^{x_i^S} \frac{du}{\delta_e(u)} - \frac{1}{2}$ is closest to an integer.

After generating the exact stop locations, the local-express system's generalized cost is recalculated. A cost model similar to Ulusoy et al.'s (2010) and Leiva et al.'s (2010) is again used for this purpose.

3.3 Numerical Analysis

Section 3.3.1 describes the demand patterns and other parameter values used in our numerical experiments. Section 3.3.2 examines an optimal local-express service design under a specific demand pattern, reporting the effectiveness of our proposed model and solution algorithm. Section 2.3.3 presents parametric analysis of local-express designs under different demand patterns.

3.3.1 Demand patterns and parameter values

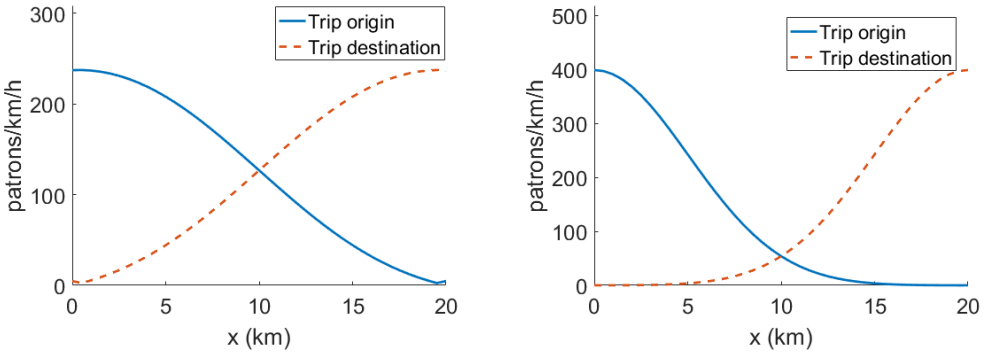
The following demand density function is considered:

$$\lambda(x, y) = \frac{1}{2}[q_1(x)q_2(y) + q_2(x)q_1(y)]\Lambda \tag{3.33}$$

where $q_1(\cdot)$ and $q_2(\cdot)$ are assumed to be truncated normal PDFs denoted by $Tr\mathcal{N}(0, \sigma^2, 0, L)$ and $Tr\mathcal{N}(L, \sigma^2, 0, L)$, respectively; i.e. normal distributions with mean 0 and L, respectively, variance σ^2, and truncated by interval $[0, L]$. The Λ is the total demand of the corridor. Note that this form of demand density function $\lambda(x, y)$ is symmetric between the two directions. Similar demand density function form was also employed in Vaughan and Cousins (1977).

In the numerical experiments presented in the following sections, it is specified that $\sigma \in \{\infty, 10, 5\}$ (km) represents three levels of demand heterogeneity, specifically: no spatial variation (uniform distribution with $\sigma = \infty$ km), low spatial variation ($\sigma = 10$ km), and high spatial variation ($\sigma = 5$ km). Figures 3.1a

and b plot the trip origin and destination densities for eastbound trips under low spatial variation ($\sigma = 10$ km) and high spatial variation ($\sigma = 5$ km), respectively, in a 20-km corridor. The average trip density for each figure is $\frac{\Lambda}{L} = 250$ trips/h/km, and the average trip distance for the three levels of demand heterogeneity is 6.7 km ($\sigma = \infty$ km), 7.6 km ($\sigma = 10$ km) and 12.1 km ($\sigma = 5$ km), respectively. The demand curves in the westbound direction are symmetric to the eastbound.



(a) Low-spatially-varied demand ($\sigma = 10$ km)

(b) High-spatially-varied demand ($\boldsymbol{\sigma = 5}$ km)

Figure 3.1 Trip origin and destination densities along the corridor for the eastbound direction

Three typical transit modes in a 20-km corridor are considered in the following sections: bus, Bus Rapid Transit (BRT), and rail. Table 3.1 presents these three transit modes' costs and operational parameters. Parameter values of $\pi_v, \pi_m, \pi_i, \pi_s, v, K$ and H_{min} for bus and rail systems are the same as in Table 2.1. Three single-mode local-express systems are examined: bus-bus, BRT-BRT, and rail-rail. In each of these systems the local and express lines are both served by the same transit mode. In addition, two bimodal local-express systems, namely, bus-BRT and bus-rail, with bus always serving local lines. We examine the average demand density $\frac{\Lambda}{L}$ varying between 100 trips/h/km and 1100 trips/h/km; and two values of time: $\mu = 5$ \$/h for a low-wage city, and $\mu = 20$ \$/h for a high-wage city. A low walking speed ($v_w = 2$ km/h) is assumed to account for signal delays and the inconvenience of walking; C_t is assumed to be 1 min/transfer ($C_t = 1.5$ min/transfer in the bus-rail system).

For our solution algorithm, the discretization interval Δx is 0.5 km, and the convergence tolerance $\varepsilon = 0.001$.

Table 3.1 Cost and operational parameters for bus, BRT and rail

	π_v (\$/veh·km)	π_m (\$/veh·h)	π_i (\$/km/h)	π_s (\$/stop/h)	τ^0 (sec)
BRT	0.66	3.81+4μ	162+5.4μ	4.2+0.14μ	30
	τ^b (sec)	τ^a (sec)	v (km/h)	K (patrons/veh)	H_{min} (min)
BRT	1	1	40	150	1

	τ^b (sec)	τ^a (sec)	τ^0 (sec)
Bus	2	1	30
Rail	0	0	45

3.3.2 Optimal design of the local-express service

The optimal design for a high-wage city ($\mu = 20$ \$/h) with a high spatially-varied demand pattern ($\sigma = 5$ km) and a demand density ($\frac{\Lambda}{L} = 250$ trips/h/km) is examined first. The optimal design results are summarized in Table 3.2. For comparison, all-stop service for the three transit modes (namely all-stop-bus, all-stop-BRT, and all-stop-rail) is included in the analysis. Note that BRT-BRT and rail-rail systems have also been tested but are not reported here. These two systems need an additional bus lane or rail track, and thus their costs appear to be always higher than other systems.

Table 3.2 illustrates that: i) for the single-mode bus system, the local-express service outperforms the conventional all-stop service by saving $100\% \times (1 - 50.29/52.69) = 4.6\%$ system cost. ii) the bimodal bus-BRT system also performs better than the all-stop-bus system (producing even higher cost saving, i.e., $100\% \times (1 - 47.44/52.69) = 10.0\%$), but its performance is only slightly superior to the all-stop-BRT system. iii) the bus-rail system cannot compete with the all-stop-bus system due to high agency costs arising from the construction and operation of rail transit; however, a comparison of bus-rail system with all-stop-rail system shows savings in both user and agency costs. In addition, the user cost saving in bimodal systems (bus-BRT and bus-rail) is huge.

Note in Table 3.2 that cost metrics are computed under the converted real design (see Section 3.2.2). Examination shows that estimation errors between CA

model results and that of real design remain less than 1% for all the numerical scenarios (including those in the following sections).

Table 3.2 System characteristics under the optimal design solution

Modes	All-stop bus	All-stop BRT	All-stop rail	Bus-bus	Bus-BRT	Bus-rail
H_e (min)	/	2.00	4.83	1.89	2.00	4.55
H_l (min)	1.82	/	/	2.67	2.76	2.53
Avg. δ_e(#/km)	/	1.78	1.02	0.60	0.70	0.40
Avg. δ_l(#/km)	1.84	/	/	3.50	3.80	3.20
User cost, (min/patron)	49.01	37.87	36.51	44.14	35.27	31.58
Agency cost (min/patron)	3.68	9.71	32.18	6.15	12.17	30.99
Generalized cost (min/patron)	52.69	47.58	68.68	50.29	47.44	64.20

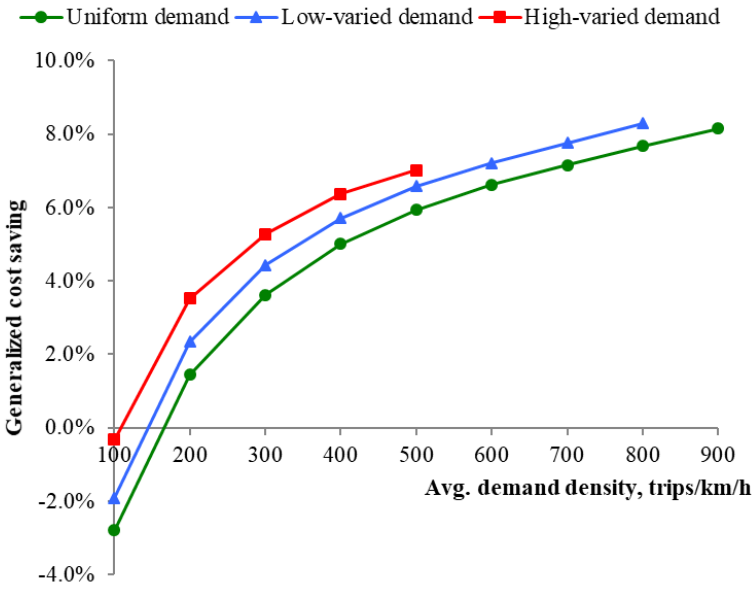
3.3.3 Parametric analysis

In this section, parametric analysis of the local-express design is conducted under three levels of demand heterogeneity (i.e. $\sigma \in \{\infty, 10, 5\}$ km), three transit systems (i.e. bus-bus, bus-BRT, bus-rail), two time values (i.e. $\mu = 5$ \$/h and $\mu = 20$ \$/h), and various average demand densities: $\frac{\Lambda}{L} \in \{100, 200, \ldots, 1100\}$ trips/h/km. For comparison, the costs of all-stop service for bus, BRT and rail system are also computed.

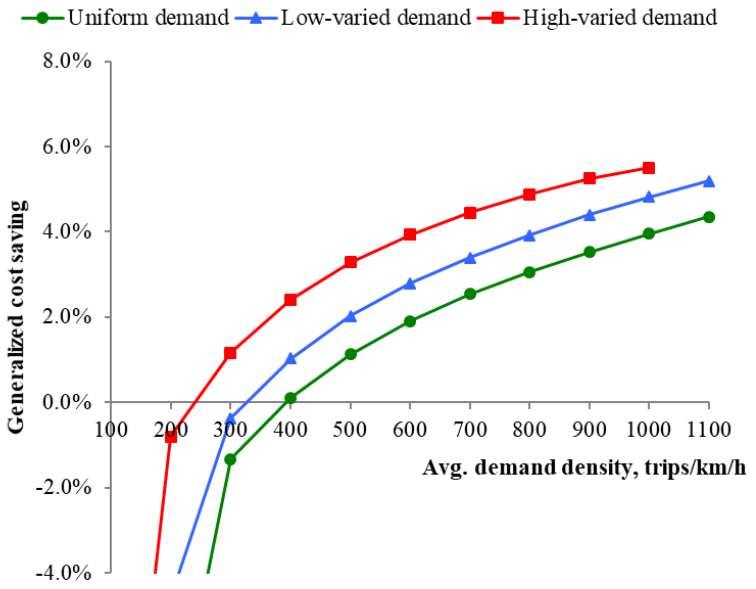
Figures 3.2a-c depict the cost savings of local-express service against all-stop service for the three transit systems in a high-wage city. Each figure shows three curves of cost saving against demand density. Specifically, the green curves with dotted markers, blue curves with triangular markers, and red curves with square markers represent the cost savings under uniform, low-spatially-varied, and high-spatially-varied demands, respectively. The cost savings are gauged by $\left(1 - \frac{SC_{bus-bus}}{AC_{bus}}\right) \times 100\%$ for the bus-bus system, $\left(1 - \frac{SC_{bus-BRT}}{\min\{AC_{bus}, AC_{BRT}\}}\right) \times 100\%$ for the bus-BRT system, and $\left(1 - \frac{SC_{bus-rail}}{\min\{AC_{bus}, AC_{rail}\}}\right) \times 100\%$ for the bus-rail system, where $SC_{bus-bus}$, $SC_{bus-BRT}$, $SC_{bus-rail}$, AC_{bus}, AC_{BRT}, and AC_{rail} are

the generalized costs for bus-bus, bus-BRT, bus-rail, all-stop-bus, all-stop-BRT, and all-stop-rail systems, respectively.

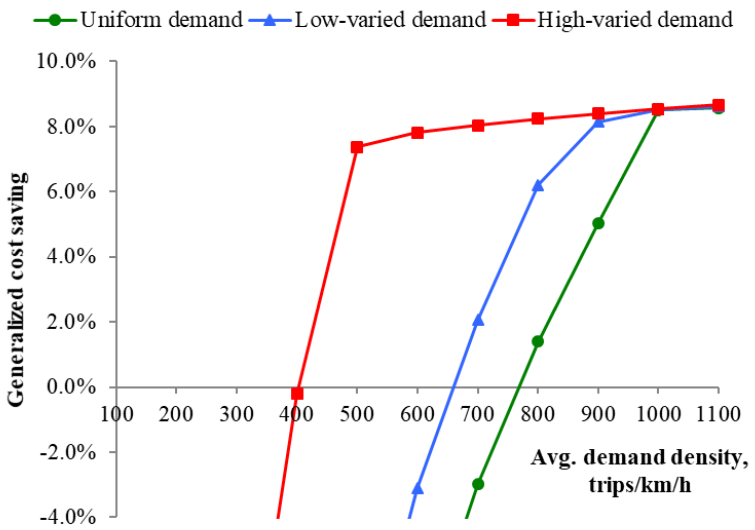
As we can see in Figure 3.2a, bus-bus system outperforms all-stop-bus system when the average demand density is greater than 100 trips/km/h. The cost saving grows with the demand density. Comparison between the three curves indicates that higher demand heterogeneity entails greater cost savings. Note that the red and blue curves end earlier than the green one since the capacity constraint is violated for the all-stop-bus system when demand is high.



(a) bus-bus versus all-stop-bus



(b) bus-BRT versus all-stop-bus and all-stop-BRT

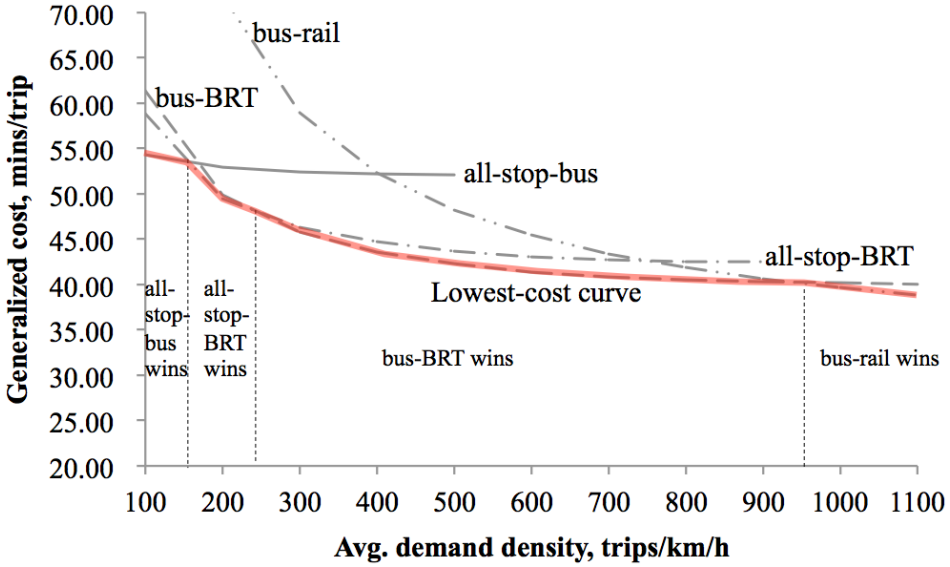


(c) bus-rail versus all-stop-bus and all-stop-rail

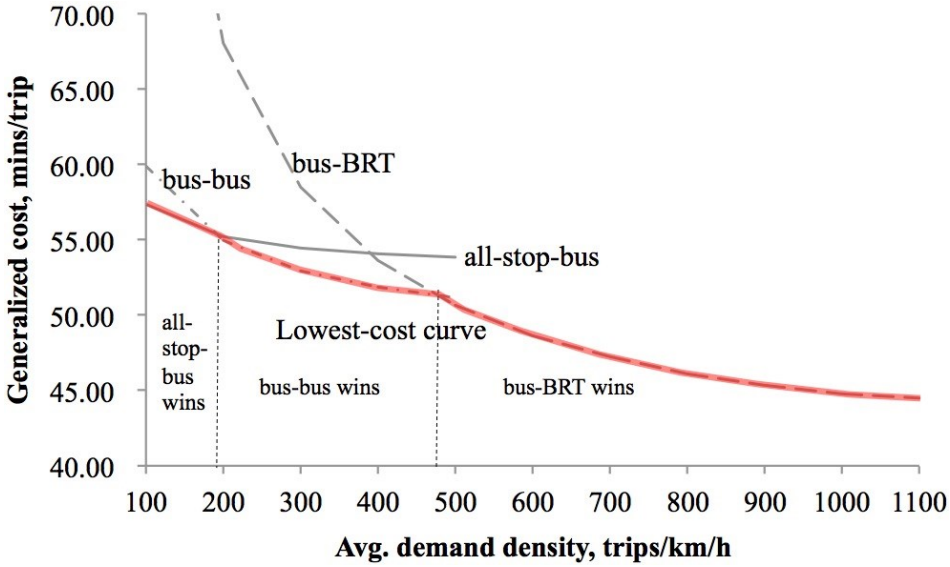
Figure 3.2 Comparison results of different systems in a high-wage city ($\mu =$ 20$/h)

Figure 3.2b shows the cost saving of bus-BRT system against all-stop-bus system or all-stop-BRT system, whichever has the lower cost. When demand density is low, the all-stop-bus system has the lowest cost among the three systems. But the bimodal system outperforms single-mode all-stop systems for high-varied demand with density greater than 300 trips/km/h, and for low-varied and uniform demands with density greater than 400 trips/km/h. It is also observed that the cost savings increase with demand density and heterogeneity.

Similar findings are observed for bus-rail system in Figure 3.2c. Since rail system has very high agency cost, the threshold demand densities for the bimodal system to win are much higher than those in the cases of bus-bus and bus-BRT systems. Furthermore, the cost savings in Figure 3.2c are less sensitive to demand density and heterogeneity.



(a) In a high-wage city ($\mu = 20$\$/h)



(b) In a low-wage city ($\mu = 5$\$/h)

Figure 3.3 The lowest-cost transit systems under high-varied demand

Figure 3.3a-b compare the costs of different all-stop and local-express designs under high-varied demand in high- and low-wage cities, respectively. The thin solid, dashed, and dotted curves in either figure illustrate the generalized costs for different systems. The red thick curve in either figure represent the lower envelope of the thin curves, highlighting the lowest-cost systems under various demand densities. In a high-wage city shown by Figure 3.3a, bimodal systems (bus-BRT and bus-rail) win for medium to high demand densities, while all-stop systems (bus and BRT) triumph only for small demand densities. For a low-wage

city shown in Figure 3.3b, a similar trend is observed, except that now bus-bus system wins for medium demands, and bus-BRT system wins for high demands.

3.4 Summary of the Local-Express Design

This chapter presents CA models for optimizing local-express service design (including both the stop locations and the routing plan) in linear corridors under arbitrary heterogeneous demand. The design allows the stop spacing of local lines and express lines to vary over the corridor, and the transit mode can be different between the local and express lines. With an embedded route assignment process, our proposed model can explicitly account for patrons' choices between different route options.

The CA model was solved by partially decomposing the formulation by the local spatial coordinate$x$, and optimizing the decision functions for each x. The undecomposable parts and the route assignment equalibrium are processed through iterative methods. Consequently a three-stage iterative algorithm is proposed to solve the model and also a discretization recipe is put forward to convert the CA solution to a real local-express design.

The superiority of local-express service over all-stop service is demonstrated by a large number of case studies. Results reveal that, compared to all-stop service, local-express service can ideally achieve generalized cost savings of over 8%. It is also shown that with the rise of demand densities and heterogeneity, bimodal local-express (bus-BRT and bus-rail) systems bring more benefits than single-mode systems.

Chapter 4 A More General Skip-Stop Design

This chapter presents the optimal design of a more general skip-stop service. In this general service different routes' non-transfer stops can be distributed arbitrarily along the line. A novel modelling method for optimizing the joint design of stop locations and routing plan has been developed, combining certain features of both discrete models and CA models. This novel model uses a discrete OD matrix as input, and specifies the stop density functions instead of the individual stop locations as decision variables. A heuristic algorithm is proposed to solve the problem. A relaxed problem has also been developed: i) to provide an initial solution for the heuristic algorithm; and ii) to furnish a lower bound for assessing the optimality gap of the heuristic solution. A discretization recipe for determining the exact stop locations is also provided. Numerical analysis verifies that this model can generate near-optimal designs serving various demand patterns. Our proposed general skip-stop model is compared with other corridor design models via a number of numerical cases and a real-world case.

Section 4.1 presents the general skip-stop design's formulation. Section 4.2 describes the proposed model's solution method. Section 4.3 presents numerical analysis. Section 4.4 summaries the general skip-stop design model.

4.1 Formulation

A bi-directional transit service operating along a linear corridor of length L is considered as shown in Figure 1.3. Demand is assumed to be exogenous. The corridor direction is denoted by $d \in \{E, W\}$, where E and W represent eastbound and westbound travel, respectively.

The formulation presented in this chapter requires to specify the number of skip-stop routes. For simplicity, only the model for two routes, numbered route A and route B, is presented. Corridors with more than two skip-stop routes can be modeled in a similar way.

For a two-route system, there are three types of stops: i) A-type stops, visited by route A only; ii) B-type stops, visited by route B only; and iii) transfer

stops (T-type stops), which are visited by both routes. These stop types are illustrated in Figure 1.3.

To facilitate modeling work, the following assumptions commonly used in the literature (Freyss et al., 2013; Gu et al., 2016) have been adopted: i) a patron always accesses or egresses the transit system through the nearest stop to her origin or destination, regardless of stop type; ii) a transit vehicle spends a constant time, τ (h), at each stop for loading and unloading patrons, including the time lost to the vehicle's acceleration and deceleration processes; iii) patrons arrive uniformly at their origin stops and always take the first vehicle that arrives; iv) a patron transfers only if there is no direct service from origin to destination, and she always transfers at the first T-type stop along the trip; and v) the two skip-stop routes' vehicle schedules are not coordinated at T-type stops.

Demand is represented by the conventional origin-destination (OD) matrix ($m \times m$) that can be obtained from a transit demand survey. This demand matrix implies that the corridor is divided into m continuous segments numbered from east to west by $k \in \{1, 2, \dots, m\}$, each with length l_k ($\sum_{k=1}^{m} l_k = L$). The $\lambda_{i,j}$ is denoted as the total demand (patrons/h) from segment i to j ($i, j \in \{1,2, \dots, m\}$). The demand is assumed to be uniformly distributed within each segment[7]. Note that when m and l_k ($k \in \{1, 2, \dots, m\}$) take different values, this demand set up can approximate any discrete or continuous demand (density) functions in the real world. Decision variables of the skip-stop service design problem include: the headways for route A and route B, denoted by H_1 and H_2, respectively; the number of stops in segment k, denoted by n_k (can be a fraction); and the proportions of A-, B-, and T-type stops in segment k, denoted by $\phi_{k,1}, \phi_{k,2}, t_k$, respectively. Our objective is to minimize the transit corridor's generalized cost, consisting of the patrons' trip cost and the agency cost. The detailed formulations are given below.

[7] A demand concentration point (e.g. the entrance to a residential community or a large shopping mall) can be represented by a very small segment.

4.1.1 User cost metrics

The patrons' cost includes four components (Daganzo, 2010): i) the access/egress cost, UC_a; ii) the waiting cost at the origin and transfer stops, UC_w; iii) the in-vehicle travel cost, UC_i; and iv) the transfer penalty, UC_t.

4.1.1.1 Access and egress cost

For segment k in the corridor, the access and egress cost equals the total walking time per hour in the segment. The average walking time for either a patron's access or egress travel is $\frac{l_k}{4n_k v_w}$, where v_w is the patrons' walking speed. The total access and egress cost in the segment is then given by:

$$\sum_{j=1}^{m} \lambda_{k,j} \cdot \frac{l_k}{4n_k v_w} + \sum_{i=1}^{m} \lambda_{i,k} \cdot \frac{l_k}{4n_k v_w}$$

Thus, the total access and egress cost of the corridor is:

$$UC_a = 2\sum_{k=1}^{m} \left(\frac{l_k}{4v_w n_k} \sum_{j=1}^{m} \lambda_{k,j} \right) \tag{4.1}$$

4.1.1.2 Waiting cost at origin and transfer stops

Waiting cost is the expected waiting time per hour for all patrons, including the waiting times at the origin stops and during transfers at transfer stops.

Specifically, a patron's waiting time depends on her trip type, which is defined by the types of her origin and destination stops. Thus there are totally nine trip types: A-type to A-type, A-type to B-type, A-type to T-type, B-type to A-type, B-type to B-type, B-type to T-type, T-type to A-type, T-type to B-type, and T-type to T-type. For each trip type, the average wait time per trip (including time spent at both origin and transfer stops) can be easily estimated. The probability of each trip type can be calculated using the proportions of the A-, B- and T-type stops in each segment. The results for all the nine trip types are summarized in Table 4.1.

Table 4.1 The nine trip types' probabilities and average wait times per trip

Trip type	Origin stop type	Destination stop type	Probability	Average wait time per trip
1	A	A	$\phi_{i,1}\phi_{j,1}$	$H_1/2$
2	A	B	$\phi_{i,1}\phi_{j,2}$	$(H_1+H_2)/2$
3	A	T	$\phi_{i,1}t_j$	$H_1/2$
4	B	A	$\phi_{i,2}\phi_{j,1}$	$(H_1+H_2)/2$
5	B	B	$\phi_{i,2}\phi_{j,2}$	$H_2/2$
6	B	T	$\phi_{i,2}t_j$	$H_2/2$
7	T	A	$t_i\phi_{j,1}$	$H_1/2$
8	T	B	$t_i\phi_{j,2}$	$H_2/2$
9	T	T	$t_i t_j$	$(H_1^{-1}+H_2^{-1})^{-1}/2$

Therefore, the average wait cost for a patron traveling from segment i to segment j, $w_{i,j}$, and the total wait cost for all the patrons, UC_w, are expressed by:

$$w_{i,j} = \frac{1}{2}\left[\sum_{r=1}^{R}\left(\left(\phi_{i,r}+\phi_{j,r}-\phi_{i,r}\phi_{j,r}\right)H_r\right)+t_i t_j H\right] \quad (4.2)$$

$$UC_w = \sum_{i=1}^{m}\sum_{j=1}^{m} w_{i,j}\,\lambda_{i,j} \quad (4.3)$$

where r is the index of the route; R is the number of routes ($R = 2$ for a two-route corridor); and $H = (\sum_{r=1}^{R} H_r^{-1})^{-1}$ is the average vehicle headway at transfer stops. Note that the real headways between vehicles at transfer stops are different from this average headway, because the speeds of different routes over the same corridor line segment are different due to the asymmetric distribution of non-transfer stops between the routes. These real headways are difficult to estimate. However, using the average headway defined above will not significantly affect the model accuracy.

4.1.1.3 In-vehicle travel cost

In-vehicle travel cost also depends on trip type. Short-distance trips of types 2 and 4 (see Table 4.1), whose origin and destination are contained between two consecutive transfer stops, must involve backtracking. In the following the cost metrics for trips without backtracking and trips involving backtracking are derived separately.

For trips without backtracking, the average on-board patron flows of route A and route B in segment k for direction $d \in \{E, W\}$ are specified as $\bar{o}_{k,1}^{d}$ and $\bar{o}_{k,2}^{d}$, respectively, and thus the average on-board patron flow in segment k for direction d is $\bar{o}_{k}^{d} = \sum_{r=1}^{R} \bar{o}_{k,r}^{d}$. We show the derivation of $\bar{o}_{k,1}^{d}$ only since $\bar{o}_{k,2}^{d}$ can be similarly derived. For a trip from segment i to segment j, the probability that the trip is completed via route A *without transfer* is $\phi_{i,1}(\phi_{j,1} + t_j) + t_i\phi_{j,1} + t_i t_j \cdot \frac{H_1^{-1}}{H^{-1}}$, where $\frac{H_1^{-1}}{H^{-1}}$ yields the probability that a patron of a T-type to T-type trip boards a route-A vehicle.

For trips that involve transfers (A-type to B-type or vice versa), the on-board patron flow will be distributed between the two routes. Since it is assumed that patrons always transfer at their first encountered T-type stop, trips traveling from B-type stops to A-type stops will mainly contribute to route A's on-board patron flow, and trips traveling from A-type stops to B-type stops will mainly contribute to route B's on-board flow. Thus, the probability that a trip from segment i to segment j is (mainly) completed via route A *with a transfer* is $\phi_{i,2}\phi_{j,1}$.

Therefore, the probability that a trip is completed via route A ($r = 1$), can be summarized as:

$$p_{i,j,r} = \phi_{j,r} + \phi_{i,r}t_j + t_i t_j \cdot \frac{H_r^{-1}}{H^{-1}} \tag{4.4}$$

Note that the on-board patron flow varies spatially within any specific segment. For simplicity, $\bar{o}_{k,1}^{d}$ is approximated by the average of the on-board flows at the left and right ends of the segment; i.e., $\bar{o}_{k,1}^{d} \approx (o_{k,1}^{d} + o_{k-1,1}^{d})/2$, where $o_{k,1}^{d}$ is the on-board flow of route A at the right end of segment k for direction d, which is defined as:

$$o_{k,1}^{E} = \sum_{i=1}^{k}\sum_{j=k+1}^{m} p_{i,j,1}\lambda_{i,j}, \forall k \in \{1, 2, \dots, m-1\};$$

$$o_{k,1}^{W} = \sum_{i=k+1}^{m}\sum_{j=1}^{k} p_{i,j,1}\lambda_{i,j}, \forall k \in \{1, 2, \dots, m-1\};$$

$$o_{k,1}^{d} = 0, \ \ \forall k \in \{0, m\}, d \in \{E, W\}.$$

To simplify the notation, a series of supplementary OD matrices $\boldsymbol{U_k^E}$ and $\boldsymbol{U_k^W}$ is introduced:

$$\boldsymbol{U_k^E} = \begin{bmatrix} 0 & \dots & 0 & \lambda_{1,k+1} & \dots & \lambda_{1,m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & 0 & \lambda_{k,k+1} & \dots & \lambda_{k,m} \\ \dots & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & \dots & 0 \end{bmatrix}, \quad \forall k \in \{1, 2, \dots, m-1\};$$

$$\boldsymbol{U_k^W} = \begin{bmatrix} 0 & \dots & \dots & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & \dots & \dots \\ \lambda_{k+1,1} & \dots & \lambda_{k+1,k} & 0 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \lambda_{m,1} & \dots & \lambda_{m,k} & 0 & \dots & 0 \end{bmatrix}, \quad \forall k \in \{1, 2, \dots, m-1\};$$

$$\boldsymbol{U_k^d} = \mathbf{0}, \ \ \forall k \in \{0, m\}, d \in \{E, W\}.$$

where each $\boldsymbol{U_k^d}$ is a $m \times m$ matrix, and the element of $\boldsymbol{U_k^d}$ at row i and column j is represented by $u_{i,j}^{k,d}$.

Thus, the on-board flow at the right end and the average on-board flow of route A ($r = 1$) for direction $d \in \{E, W\}$ can be reformulated as:

$$o_{k,r}^d = \sum_{i=1}^m \sum_{j=1}^m p_{i,j,r}\, u_{i,j}^{k,d}, \forall\, k \in \{1, 2, \dots, m\}, d \in \{E, W\} \quad (4.5a)$$

$$\bar{o}_{k,r}^d = \frac{\left(o_{k,r}^d + o_{k-1,r}^d\right)}{2} = \sum_{i=1}^m \sum_{j=1}^m p_{i,j,r}\, \bar{u}_{i,j}^{k,d}, \forall k \in \{1, 2, \dots, m\}, d \in \{E, W\} \quad (4.5b)$$

where $\bar{u}_{i,j}^{k,d} = \left(u_{i,j}^{k,d} + u_{i,j}^{k-1,d}\right)/2$. Note that the total average on-board flow in segment k for direction d, i.e. $\bar{o}_k^d = \sum_{r=1}^R \bar{o}_{k,r}^d = \sum_{i=1}^m \sum_{j=1}^m \bar{u}_{i,j}^{k,d}$, is a constant value for a given demand pattern.

Consequently, the cost of trips without backtracking for segment k can be expressed as

$$\Gamma_k = \sum_{r=1}^R \left(\bar{o}_{k,r}^E + \bar{o}_{k,r}^W\right) \left[\frac{l_k}{v} + \tau\left(\phi_{k,r} + t_k\right) n_k\right] \quad (4.6)$$

where v is the vehicle cruise speed (assumed to be indifferent between route A and B); and $\left(\phi_{k,r} + t_k\right) n_k$ yields the total number of stops in segment k for route r.

For backtracking trips within the skip-stop bays in segment k, the number of transfer stops is $t_k n_k$ and the stop spacing between two transfer stops can be approximated by $s_k = \min\left(L, x_k + \frac{l_k}{2 t_k n_k}\right) - \max\left(0, x_k - \frac{l_k}{2 t_k n_k}\right)$, where x_k is

the coordinate of segment k's midpoint. Note that s_k may span more than one segment if $t_k n_k < 1$. Hence, we denote the indices of the left-most and right-most segments spanned by s_k as bl_k and br_k, respectively. These are given by:

$$bl_k = I\left(\max\left(0, x_k - \frac{l_k}{2t_k n_k}\right)\right) \tag{4.7a}$$

$$br_k = I\left(\min\left(L, x_k + \frac{l_k}{2t_k n_k}\right)\right) \tag{4.7b}$$

where $I(x)$ is an index function that maps coordinate x to the index of the segment containing x.

We can now approximate the total number of backtracking trips within s_k by summing up all the contained trips involving transfers. This is given by $\sum_{i=bl_k}^{br_k}\sum_{j=bl_k}^{br_k}\hat{p}_{i,j}\lambda_{i,j}$, where $\hat{p}_{i,j}$ is the probability that a trip involves a transfer, which is written as:

$$\hat{p}_{i,j} = \sum_{r=1}^{R}\phi_{i,r}(1 - \phi_{j.r} - t_j) \tag{4.8}$$

Then, we calculate the average backtracking density for segment k, b_k (trips per km), as follows:

$$b_k = \frac{\sum_{i=bl_k}^{br_k}\sum_{j=bl_k}^{br_k}\hat{p}_{i,j}\lambda_{i,j}}{\sum_{s=bl_k}^{s=br_k} l_s} \tag{4.9}$$

Since a skip-stop bay is not long (Freyss et al., 2013), we here approximate the backtracking cost by assuming that those trips are uniformly distributed in the skip-stop bay. Under this assumption, the total backtracking vehicle distance is $\frac{1}{3}s_k b_k$, where $\frac{1}{3}s_k$ is the average backtracking distance per trip in segment k (see the detailed derivation in Gu, et al., 2016). Recognizing that half the backtracking flow is on route A and the other half is on route B, the additional cost incurred by backtracking trips in segment k can be expressed as $\Theta_k = \left(\frac{l_k}{v} + \frac{\tau}{2}n_k\right)\frac{1}{3}s_k b_k$. As s_k is a non-differentiable function of t_k and n_k, it is approximated by a tight upper bound, $\frac{\alpha \cdot l_k}{t_k n_k + \frac{1}{L}}$, where $\alpha = \max_k \frac{l_k + 1}{l_k}$ (see Appendix D for the derivation of this upper bound). The additional backtracking cost is thus re-written as:

$$\Theta_k = \frac{\alpha}{3}\left(\frac{l_k}{v} + \frac{\tau}{2}n_k\right)\frac{l_k}{t_k n_k + \frac{1}{L}}b_k \tag{4.10}$$

This approximation does slightly over-estimate the backtracking cost, but it is conservative and greatly simplifies the problem solving.

Combining the costs of trips with and without backtracking, the total in-vehicle travel cost can be calculated as

$$UC_i = \sum_{k=1}^{m} [\Gamma_k + \Theta_k] \tag{4.11}$$

where the costs of trips without and with backtracking, Γ_k and Θ_k, are given by equations (4.4-4.10).

4.1.1.4 Transfer penalty

An extra cost penalty for transfers is considered, which is expressed by

$$UC_t = \sum_{i=1}^{m} \sum_{j=1}^{m} C_t \hat{p}_{i,j} \lambda_{i,j} \tag{4.12}$$

where C_t is the unit transfer penalty (h/transfer); and $\hat{p}_{i,j}$ is given by equation (4.8).

4.1.2 Agency cost metrics

The agency cost consists of: i) the distance-based vehicle operating cost (for example fuel costs), AC_K; ii) the time-based vehicle operating cost (such as amortized vehicle purchase cost and staff wages), AC_H; iii) the amortized line infrastructure cost (such as building costs for busway or rail tracks), AC_I; and iv) the amortized stop infrastructure cost, AC_S. These are formulated as follows:

$$AC_K = \frac{2\pi_v L}{\mu} \sum_{r=1}^{R} H_r^{-1} \tag{4.13}$$

$$AC_H = \frac{2\pi_m}{\mu} \sum_{k=1}^{m} \sum_{r=1}^{R} \left(\frac{l_k}{v} + \tau(\phi_{k,r} + t_k) n_k \right) H_r^{-1} \tag{4.14}$$

$$AC_I = \frac{2\pi_i L}{\mu} \tag{4.15}$$

$$AC_S = \frac{\pi_s}{\mu} \sum_{k=1}^{m} n_k \tag{4.16}$$

where μ is the value of time (\$/h); π_v and π_m are the unit operating costs per vehicle-hour and per vehicle-km, respectively; π_i and π_s are the unit infrastructure construction and maintenance costs per km of line infrastructure and per stop, respectively, amortized for each service hour.

4.1.3 Problem formulation

The generalized cost, denoted by SC, is the sum of user and agency costs. Those cost terms have been introduced by equations (4.1-4.16) for a two-route skip-stop service model, i.e. $R = 2$. Without further modifications, the above equations are still valid for more general multi-route cases, i.e. $R \geq 3$, Thus, the cost minimization problem can be formulated as:

$$\min SC = UC_a + UC_w + UC_i + UC_t + AC_K + AC_H + AC_I + AC_S \quad (4.17a)$$

subject to:

$$H_{min} \leq H_1 \leq H_2 \ldots \leq H_R \leq H_{max} \quad (4.17b)$$

$$\frac{K}{H_r} \geq \bar{o}_{k,r}^{d} + B_k, \quad \forall r \in \{1,2,\ldots,R\}, k \in \{1,2,\ldots,m\}, d \in \{E,W\} \quad (4.17c)$$

$$0 \leq \phi_{k,r} \leq 1, \quad \forall r \in \{1,2,\ldots,R\}, k \in \{1,2,\ldots,m\} \quad (4.17d)$$

$$0 \leq t_k \leq 1, \quad \forall k \in \{1,2,\ldots,m\} \quad (4.17e)$$

$$\sum_{r=1}^{R} \phi_{k,r} + t_k = 1, \forall k \in \{1,2,\ldots,m\} \quad (4.17f)$$

$$0 \leq n_k \leq \bar{N}_k = \frac{l_k}{s_{min}}, \quad \forall k \in \{1,2,\ldots,m\} \quad (4.17g)$$

where H_{min} and H_{max} are the minimum and maximum dispatching headways for each skip-stop route, respectively; K is the vehicle's passenger-carrying capacity; B_k is the maximum backtracking flow in segment k in one direction, which is approximately equal to half of the total backtracking flow in the segment; $B_k = \frac{\alpha}{2t_k n_k + \frac{2}{L}} b_k$. This assumes the trips involving backtracking are uniformly distributed within the segment; see Gu et al. (2016) for more details on this issue. Note that to break the symmetry between the routes, we specify that $H_{r+1} \geq H_r, \forall r \in \{1,..,R-1\}$ in constraint (4.17b). Constraints (4.17c) ensure the number of patrons onboard a transit vehicle never exceeds the vehicle capacity; note that the RHS of the inequality is the on-board flow of segment k for direction $d \in \{E,W\}$. Constraints (4.17-f) specify that the probability of different stop types is between 0 and 1 and their probabilities are summed up to 1 in each segment. Constraints (4.17g) stipulate that the number of stops in each segment must be positive and stop spacing must be greater than a minimum stop spacing of s_{min}.

4.2 Solution Method

Section 4.2.1 presents a heuristic algorithm to solve the model. Section 4.2.2 describes a relaxed problem to provide: i) an initial solution to the heuristics, and ii) a lower bound. Section 4.2.3 introduces a recipe to convert the model solution to a design with exact stop locations.

4.2.1 Heuristic solution method

The decision variables in problem (4.17) include R variables of H_r, $R \times m$ variables of $\phi_{k,r}$, m variables of t_k, and m variables of n_k. It should be noted that the number of routes R is typically small ($R \leq 4$), since a system with more than 4 routes is difficult to operate in real practice. A heuristic algorithm to solve (4.17) for a given R is presented next.

A two-stage algorithm is proposed to solve the problem. In the first stage, we fix H_r, and optimize (4.17a,4.17c-g) with respect to $\phi_{k,r}, t_k$ and n_k. Note that b_k in objective function (4.17a) and constraint (4.17c) involves an indicator function, which is discontinuous and indifferentiable; see (4.7-4.9). An iterative method is thereby employed to alternately evaluate b_k and optimize $\phi_{k,r}, t_k$ and n_k. In the nth iteration, $\phi_{k,r}^{(n)}, t_k^{(n)}, n_k^{(n)}$ are optimized by solving the following non-linear program for given $b_k = b_k^{(n-1)}$:

$$\left(\phi_{k,r}^{(n)}, t_k^{(n)}, n_k^{(n)}\right) = \min_{\phi_{k,r}, t_k, n_k} SC|H_r, b_k^{(n-1)}$$

subject to constraints (4.17c-g). (4.18)

The non-linear program is solved via a standard sequential quadratic programming (SQP) method. After obtaining the solution, the new values of b_k, denoted by b_k', are calculated using (4.7-4.9). The b_k is then updated via the method of successive averages (MSA) (Sheffi, 1985), given by:

$$b_k^{(n)} = b_k^{(n-1)} + \frac{b_k' - b_k^{(n-1)}}{n} \quad (4.19)$$

The iterative method is repeated until convergence is reached, thus obtaining the local optimum $\phi_{k,r}^*, t_k^*, n_k^*$ for the given H_r.

The second stage fixes $\phi_{k,r}, t_k$ and n_k and optimizes (4.17a-c) with respect to H_r. The problem addressed in this stage is also nonlinear, and it is again solved via the SQP method.

The two-stage algorithm is executed iteratively until convergence is reached. The algorithm's steps of are summarized as follows:

Step 1. Set the initial values of decision variables, $H_r^{(0)}$, $\phi_{k,r}^{(0)}$, $t_k^{(0)}$, and $n_k^{(0)}$ for all $k \in \{1,2,\dots,m\}, r \in \{1,2,\dots R\}$, and calculate $b_k^{(0)}$ using (4.7-4.9).

Step 2. For given $H_r^{(j)}$, apply the iterative method to find the local optimum $\phi_{k,r}^*, t_k^*, n_k^*$ using (4.18-4.19) until $\left|\frac{b_k^{(n)} - b_k^{(n-1)}}{b_k^{(n-1)}}\right| \leq \varepsilon$ is satisfied for each k, where ε is a pre-specified tolerance (e.g., $\varepsilon = 0.001$).

Step 3. Find the optimal H_r^* via the SQP method. Set $H_r^{(j+1)} = H_r^*$, and $j = j + 1$.

Step 4: Repeat steps (2-3) until $\sum_{r=1}^{R} \left|\frac{H_r^{(j)} - H_r^{(j-1)}}{H_r^{(j-1)}}\right| \leq \varepsilon$.

4.2.2 The relaxed problem

In the relaxed problem, some items in the objective function are reduced and some constraints related to backtracking trips are dropped. In addition, high degree polynomial terms are replaced by smaller, quadratic terms. The problem is thus reduced to a quadratic program. This relaxed problem's optimal solution, which is much easier to obtain, will thus be a lower bound of the optimum of the original problem. Furthermore, the relaxed problem retains the essential cost terms and constraints related to the proportions of non-transfer stops in each segment, and thus the underlying trade-off between different routes is still by-and-large captured. Hence, the optimal solution of the relaxed problem is also used as an initial solution for the two-stage algorithm described in Section 4.2.1.

The relaxed problem is defined for fixed H_r $(r = 1, \dots, R)$. To create this problem, the cost terms related to backtracking in the objective function (4.17a) are first removed and then the constraint (4.17b) is relaxed. The original problem

is then reduced to program (4.20a-f). Note that the minimal cost of (4.20a-f) is a lower bound of the original problem (4.17) for any given H_r.

$$\min SC1|H_r = \sum_{i=1}^{k} \frac{a_k}{n_k} + g(\boldsymbol{\phi}, \boldsymbol{t}, \boldsymbol{n}) + \tau \sum_{k=1}^{m} \sum_{r=1}^{R} (\bar{o}_{k,r}^{E} + \bar{o}_{k,r}^{W}) \phi_{k,r} n_k + C \tag{4.20a}$$

subject to:

$$\frac{K}{H_r} \geq \sum_{i=1}^{m} \sum_{j=1}^{m} \phi_{j,r}\, u_{i,j}^{k,d}\,, \forall r \in \{1, 2, \dots, R\}, k \in \{1,2, \dots, m\}, d \in \{E, W\} \tag{4.20b}$$

$$0 \leq \phi_{k,r} \leq 1, \quad \forall r \in \{1, 2, \dots, R\}, k \in \{1,2, \dots, m\} \tag{4.20c}$$

$$0 \leq t_k \leq 1, \quad \forall k \in \{1,2, \dots, m\} \tag{4.20d}$$

$$\sum_{r=1}^{R} \phi_{k,r} + t_k = 1, \forall k \in \{1,2, \dots, m\} \tag{4.20e}$$

$$0 \leq n_k \leq \bar{N}_k = \frac{l_k}{s_{min}}, \quad \forall k \in \{1,2, \dots, m\} \tag{4.20f}$$

where $a_k = \frac{l_k \sum_{j=1}^{m} \lambda_{k,j}}{2 v_w}$ is a constant coefficient of $\frac{1}{n_k}$; $g(\boldsymbol{\phi}, \boldsymbol{t}, \boldsymbol{n})$ is a quadratic function of decision variables $(\phi_{k,r}, t_k, n_k)$ for $k \in \{1,2, \dots, m\}$, $r \in \{1,2, \dots, R\}$ (the decision variables are represented by vectors $\boldsymbol{\phi}, \boldsymbol{t}$, and $\boldsymbol{n}$) given by: $g(\boldsymbol{\phi}, \boldsymbol{t}, \boldsymbol{n}) = \sum_{i=1}^{m} \sum_{j=1}^{m} w_{i,j}\, \lambda_{i,j} + \sum_{i=1}^{m} \sum_{j=1}^{m} C_t \hat{p}_{i,j} \lambda_{i,j} + \frac{2\pi_m \tau}{\mu} \sum_{k=1}^{m} \sum_{r=1}^{R} \big(\phi_{k,r} + t_k\big) n_k H_r^{-1} + \sum_{k=1}^{m} \tau (\bar{o}_k^E + \bar{o}_k^W) t_k n_k + \frac{\pi_s}{\mu} \sum_{k=1}^{m} n_k$; and C is a constant value given by $C = \sum_{i=k}^{R} \frac{(\bar{o}_k^E + \bar{o}_k^W) l_k}{v} + \frac{2\pi_v L}{\mu} H^{-1} + \frac{2\pi_m}{\mu} \frac{L}{v} H^{-1} + \frac{2\pi_i L}{\mu}$. Note that constraints (4.20b-f) are all linear constraints in problem (4.20).

The cost term $\tau \sum_{k=1}^{m} \sum_{r=1}^{R} (\bar{o}_{k,r}^{E} + \bar{o}_{k,r}^{W}) \phi_{k,r} n_k$ in (4.20a) is a fourth degree polynomial. We further reduce this cost term to a quadratic function by using the inequalities $\bar{o}_{k,r}^{d} \geq \sum_{i=1}^{m} \sum_{j=1}^{m} \Big(\phi_{j,r} + \max\{\phi_{i,r} + t_j - 1, 0\} + \frac{H_r^{-1}}{H^{-1}} \max\{t_i + t_j - 1, 0\}\Big) \bar{u}_{i,j}^{k,d}$ and $n_k \geq n_k^{min}$, where $n_k^{min} = \sqrt{a_k} \left(\frac{\pi_s}{\mu} + \tau(\bar{o}_k^E + \bar{o}_k^W) + \frac{2\pi_m \tau}{\mu} H^{-1}\right)^{-\frac{1}{2}}$. The problem is then reduced to:

$$\min SC2|H_r = \sum_{i=1}^{k} \frac{a_k}{n_k} + g(\boldsymbol{\phi}, \boldsymbol{t}, \boldsymbol{n}) + \tau \sum_{k}^{m} \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{r=1}^{m} \Big(\phi_{j,r} + \sigma_{i,j,r} + \frac{H_r^{-1}}{H^{-1}} \beta_{i,j}\Big) \phi_{k,r} (\bar{u}_{i,j}^{k,E} + \bar{u}_{i,j}^{k,W}) n_k^{min} + C \tag{4.21a}$$

subject to constraints (4.20b-e), and

$$n_k^{min} \le n_k \le \bar{N}_k = \frac{l_k}{s_{min}} \quad \forall k \in \{1,2,\dots,m\} \tag{4.21b}$$

$$\sigma_{i,j,r} \ge \phi_{i,r} + t_j - 1 \tag{4.21c}$$

$$\sigma_{i,j,r} \ge 0 \tag{4.21d}$$

$$\beta_{i,j} \ge t_i + t_j - 1 \tag{4.21e}$$

$$\beta_{i,j} \ge 0 \tag{4.21f}$$

Constraints (4.21b) imply that $n_k^{min} \le \bar{N}_k$, which is true in real practice. In the worst case, if there exists some $k' \in \{1,\dots,m\}$ such that $n_{k'}^{min} > \bar{N}_k$, the constraints can be modified to $n_{k'} = \bar{N}_{k'}$ for these k' and the following proposition still holds. For simplicity, discussion of the above special case is omitted. The following proposition shows that problem (4.21) is a lower bound problem of (4.20). Its proof is furnished in Appendix E.

Proposition 1. Assuming that there is at least a feasible solution to problem (4.21), the minimum of problem (4.21) must be a lower bound of problem (4.20).

Finally, the items $\frac{1}{n_k}$ are approximated in the objective function (4.21a) using a linear approximation method. For each segment k, we evenly select P numbers $N_k^1, N_k^2, \dots, N_k^P$ from the interval $\left[n_k^{min}, N_k\right]$, and define set $\boldsymbol{\mathcal{N}_k} = \{N_k^1, N_k^2, \dots, N_k^P\}$. A linear approximation of $\frac{1}{n_k}$ near the point $N_k^p \in \boldsymbol{\mathcal{N}_k}$ can be expressed as $\frac{1}{n_k} \approx \left(N_k^p\right)^{-1} - \left(N_k^p\right)^{-2} n_k$. Since $\frac{1}{n_k}$ is a convex function, its linear approximation in domain $\left[n_k^{min}, N_k\right]$ is given by:

$$\frac{1}{n_k} \ge \max_{\mathrm{p}} \left\{2\left(N_k^p\right)^{-1} - \left(N_k^p\right)^{-2} n_k\right\}, \ \forall N_k^p \in \boldsymbol{\mathcal{N}_k} \tag{4.22}$$

Hence, we substitute $\frac{1}{n_k}$ in problem (4.21a) by (4.22). Problem (4.21) can then be reduced to

$$\min SC3|H_r = \sum_{i=1}^{k} a_k y_k + g(\boldsymbol{\phi}, \boldsymbol{t}, \boldsymbol{n}) + \tau \sum_k^m \sum_{i=1}^m \sum_{j=1}^m \sum_{r=1}^m \left(\phi_{j,r} + \sigma_{i,j,r} + \frac{H_r^{-1}}{H^{-1}} \beta_{i,j}\right) \phi_{k,r} \left(\bar{u}_{i,j}^{k,E} + \bar{u}_{i,j}^{k,W}\right) n_k^{min} + C \tag{4.23a}$$

subject to constraints (4.20b-e), (4.21b-f), and

$$y_k \geq 2\left(N_k^p\right)^{-1} - \left(N_k^p\right)^{-2} n_k, \ \forall N_k^p \in \boldsymbol{\mathcal{N}_k} \tag{4.23b}$$

Obviously, problem (4.23) is a lower bound problem of problem (4.17). Note that problem (4.23) is a nonconvex quadratic problem (QP), which has a nonconvex quadratic objective function and linear constraints. This problem can be solved to global optimality by commercial solvers, such as CPLEX.

Note now that the remaining decision variables, H_r, can be optimized using exhaustive search with a fixed interval of for instance 0.5 min, since there are at most R variables of H_r, and these are bounded by constraint (4.17b). The resulting solution is near optimal for the relaxed problem, if the search interval of the exhaustive search is sufficiently small. This solution can then be used as an initial solution for the algorithm presented in Section 4.2.1. In addition, a lower bound on the optimal solution to the original problem can be produced, if the search interval for H_r is set to a very small value. This lower bound can be used to examine the algorithm's solution quality. However, the process for obtaining the lower bound may take a long time due to the small search interval.

4.2.3 Generating the exact stop locations

The numbers of stops for each stop type, obtained using the above algorithm, are fractions for each segment. These fractions cannot be directly applied to design a real skip-stop service. Hence, the following recipe has been developed to convert the optimal solution to a real world design with exact stop locations.

First, the number of stops in each segment is determined, denoted by $\bar{n}_k$ for segment k. It is given by

$$\bar{n}_k = \left\lfloor \sum_{u=1}^{k} n_u^* - \sum_{u=1}^{k-1} \bar{n}_u \right\rfloor, \forall k = \{2,3, \dots, m\}$$

$$\bar{n}_1 = \lfloor n_1^* \rfloor \tag{4.24}$$

where n_u^* is obtained by solving (4.17); and $\lfloor \cdot \rfloor$ is the floor function. Hence, there are $M = \sum_{k=1}^{m} \bar{n}_k$ stops located in the corridor with even stop spacing $\frac{l_k}{\bar{n}_k}$ in each segment k.

The stops for each route r are then determined. As $\phi_{k,r}$ and t_k represent the proportion of different stop types in a segment, these are re-indexed with respect to each stop, denoted by P_z^r, where $z \in \{1, \dots M\}$ is the index of stops; $r \in$

$\{0,1,\ldots,R\}$ corresponds to the type of stop and $r = 0$ represents the transfer stop. Then the stop assignment can be obtained by iteratively calculating the following two equations:

$$r_z = \arg\max_{\mathrm{r}}\{\sum_{s=1}^{z} P_s^r - C_{z-1}^r\} \tag{4.25}$$

$$C_z^r = \begin{cases} C_{z-1}^r + 1, & \text{if } r = r_z \\ C_{z-1}^r, & \text{otherwise} \end{cases} \tag{4.26}$$

where r_z represents the stop type of stop z, and C_z^r is the cumulative number of stops for stop type r, counting from stop 1 to stop z ($C_0^r \equiv 0$).

4.3 Numerical Analysis

Section 4.3.1 describes the demand patterns and other parameter values used in our numerical experiments. Section 4.3.2 reports the validation tests of our CA models. Section 4.3.3 presents parametric analysis of local-express designs under different demand patterns, and compares the general skip-stop model with other corridor design models. Section 4.3.4 examines a real-world case study.

4.3.1 Demand patterns and parameter values

A $m \times m$ demand matrix of the following form is considered:

$$\lambda_{i,j} = \left[(1-\rho)q_{i,j}^1 + \rho q_{i,j}^2\right]\Lambda \tag{4.27}$$

where Λ is the total demand of the corridor; $q_{i,j}^1$ and $q_{i,j}^2$ are assumed to be two different OD probability matrices, representing the probability of a trip originating from segment i and terminating at segment j in the corridor. The $q_{i,j}^1$ is specified as a uniform OD probability matrix, i.e. $q_{i,j}^1 = \frac{1}{m(m-1)}$ for all $i \neq j$ and $q_{i,j}^1 = 0$ for $i = j$; and $q_{i,j}^2$ is used to represent the demand heterogeneity. $\rho \in [0,1]$ is a coefficient used to adjust the proportion of the uniformly distributed demand in the total demand, i.e. $\rho = 0$ represents that $\lambda_{i,j}$ is uniformly distributed and $\rho = 1$ reflects the greatest demand heterogeneity.

Numerical studies in the following sections study a 20-km linear corridor. A "chessboard" shape probability OD matrix $q_{i,j}^2$ is specified, as illustrated in Figure 4.1, where the trips' probability in a unit area is represented by colors. The

probability that a trip is in a km^2 of yellow area in the figure is 0.52%, and the probability that a trip is in a km^2 of black area is 1.56% (one can verify that these probabilities sum up to 1 over the entire OD region). Examination of ρ from 0 to 1 is conducted to test the skip-stop model's performance under different degrees of demand heterogeneity. For validation of our model, note that the optimal design of a "chessboard" shape demand pattern ($\rho = 1$) should have two independent routes: one serving the demand in the yellow squares and the other serving the demand in the black squares. T-type stop should not exist because no patron needs to transfer. On the other hand, if the demand is uniform ($\rho = 0$), the optimal design will be symmetric (for example, an AB-type service).

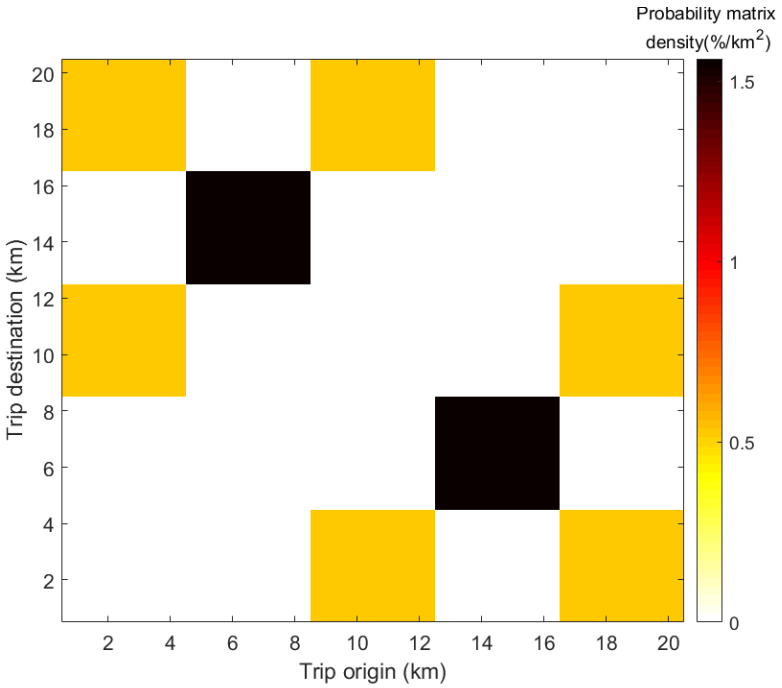


Figure 4.1 The "chessboard" OD probability matrix

Two typical transit modes, a bus system and a rail system, are examined. These two transit modes' cost and operational parameters are the same as in Table 2.1, but the unit infrastructure construction π_i is doubled for the rail system to account for the additional rail track needed for overtaking. We examine the average demand density $\frac{\Lambda}{L} \in \{100,200,300\}$ trips/h/km in a bus system, and $\frac{\Lambda}{L} \in \{500,700,900\}$ trips/h/km in a rail system; and two time values: $\mu = 5$ \$/h for a low-wage city, and $\mu = 20$ \$/h for a high-wage city. A low walking speed ($v_w =$

2 km/h) is assumed to account for signal delays and the inconvenience of walking; C_t is assumed to be 1 min/transfer.

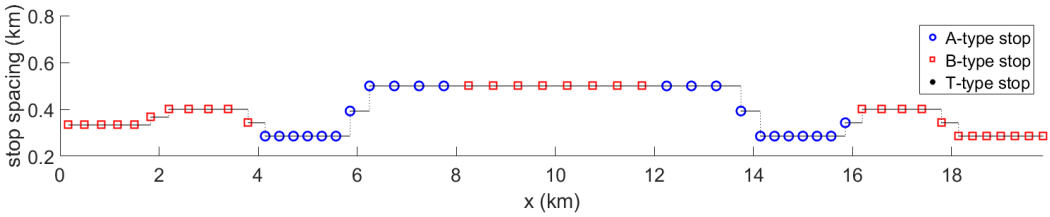
For our solution algorithm, an evenly spaced OD matrix with $m = 20$ and $l_k = 1$ km, $\forall k \in \{1,2,\dots,m\}$ is considered. The convergence tolerance is $\varepsilon = 0.001$. When finding the initial solution, H_r is searched with a step of 1 min, and the local minimum solver in CPLEX is employed to rapidly solve the quadratic programming (see Section 4.2.2).

4.3.2 Model validation

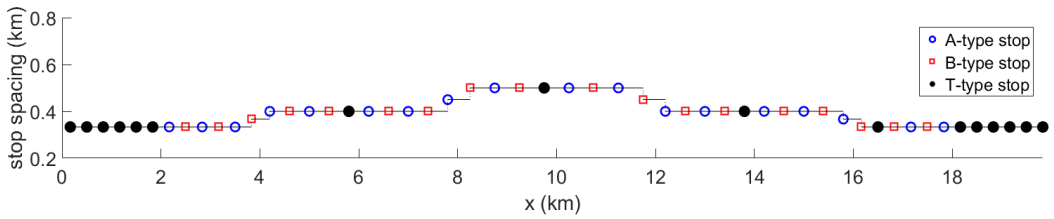
Our model is validated using two specific demand patterns in Section 4.3.2.1. Solution quality is examined in Section 4.3.2.2. Section 4.3.2.3 presents the CA solution's computational efficiency and approximation error with respect to segment length.

4.3.2.1 Optimal designs under specific demand patterns

The optimal design results are examined under the two specific demand patterns, namely the "chessboard" shape demand pattern ($\rho = 1$), and the uniform demand pattern ($\rho = 0$). Figures 4.2a-b present the optimal design results for a bus corridor under these two demand patterns in a high-wage city ($\mu = 20$ \$/h) with average demand density of $\frac{\Lambda}{L} = 200$ trips/h/km. As expected, Figure 4.2a shows a design of two independent routes, in which the blue circles represent A-type stops, and red squares represent B-type stops. No transfer stop is needed. Figure 4.2b shows an AB-type design, where the T-type stops are marked by black dots. Note that many T-type stops are located at the corridor's two ends. This is because the on-board passenger flow is low near the corridor ends, thus stop spacing of each route can be made smaller.



(a) "chessboard" shape demand ($\rho = 1$)



(b) uniform demand ($\rho = 0$)

Figure 4.2 Optimal stop spacing under different demand patterns

4.3.2.2 Solution quality

The quality of our CA solution is first compared to the lower bound obtained from (4.23). Then the approximation error of the CA model as compared with the converted design (see Section 4.2.3) is evaluated.

To find the lower bound, the relaxed problem is solved by using a small search interval of H_r (0.1 min) and applying the global quadratic programming solver in CPLEX (see Section 4.2.2). The relative cost gap between the lower bound and the CA solution is calculated for the numerical instances in Section 4.3.1. The average cost gap is found to be about 1.8% and the maximum cost gap is only 3.1%. Considering that the lower bound is obtained by ignoring some cost terms, the above cost gaps manifest good quality of our CA solution.

Approximation errors between the CA solution and the converted design for the numerical instances are summarized in Table 4.2. The table shows that the error for bus systems never exceeds 1.1% and is only 0.3% on average, and the error for rail systems never exceeds 1.8% and is 0.8% on average. Hence, our CA model is quite accurate in approximating the system's generalized cost. The approximation errors mainly arise from rounding off the number of stops in each segment. This also explains why the rail system has larger approximation errors, as it has fewer stops in each segment.

Table 4.2 Comparison between the costs of CA solution and converted design

	Bus systems		Rail systems	
	Average error (%)	Maximum error (%)	Average error (%)	Maximum error (%)
Generalized cost, SC	0.3%	1.1%	0.8%	1.8%
User cost	0.5%	1.3%	3.2%	5.2%
Agency cost	1.7%	2.2%	1.5%	1.8%
UC_a	1.1%	2.1%	2.3%	3.0%
UC_w	1.4%	4.6%	2.7%	4.9%
UC_i	0.5%	1.4%	3.8%	8.3%
UC_t	6.9%	14.9%	13.3%	24.3%
AC_K	0.6%	0.8%	1.3%	1.3%
AC_H	2.0%	2.7%	5.2%	7.2%
AC_I	0.6%	0.8%	1.3%	1.3%
AC_S	0.9%	2.0%	1.5%	2.7%

4.3.2.3 Effects of segment length

The segment length significantly affects the estimation error and the solution time. Understanding these effects is useful for guiding demand data collection and transit corridor optimization procedures.

In this section, the solution procedure is performed using the numerical instances defined in Section 4.3.1 with different segment lengths. For simplicity, four uniform segment lengths of 0.5 km, 1 km, 2 km, and 4km are used. The solution algorithm was performed in Matlab 2016a on a personal computer with Intel Core i7-4970 CPU @ 3.60 GHz and 16G RAM.

Table 4.3 presents the computation time and the percentage cost error between CA solution and converted design for the four segment lengths. It shows that the computation time increases as the segment length decreases. However, the estimation error is minimum when 1-km segments are used. Further smaller segments will only yield greater errors. This is possibly because the round-off error for the number of stops increases as the segment length decreases, since a shorter segment contains a smaller fraction of stops.

Table 4.3 Comparison between different segment lengths

Segment length	0.5 km	1 km	2 km	4 km
Computation time (s)	363.7	216.5	124.5	82.8
Average percentage cost error between CA solution and converted design	0.8%	0.3%	1.5%	4.4%
Maximum percentage cost error between CA solution and converted design	2.3%	1.1%	3.3%	9.1%

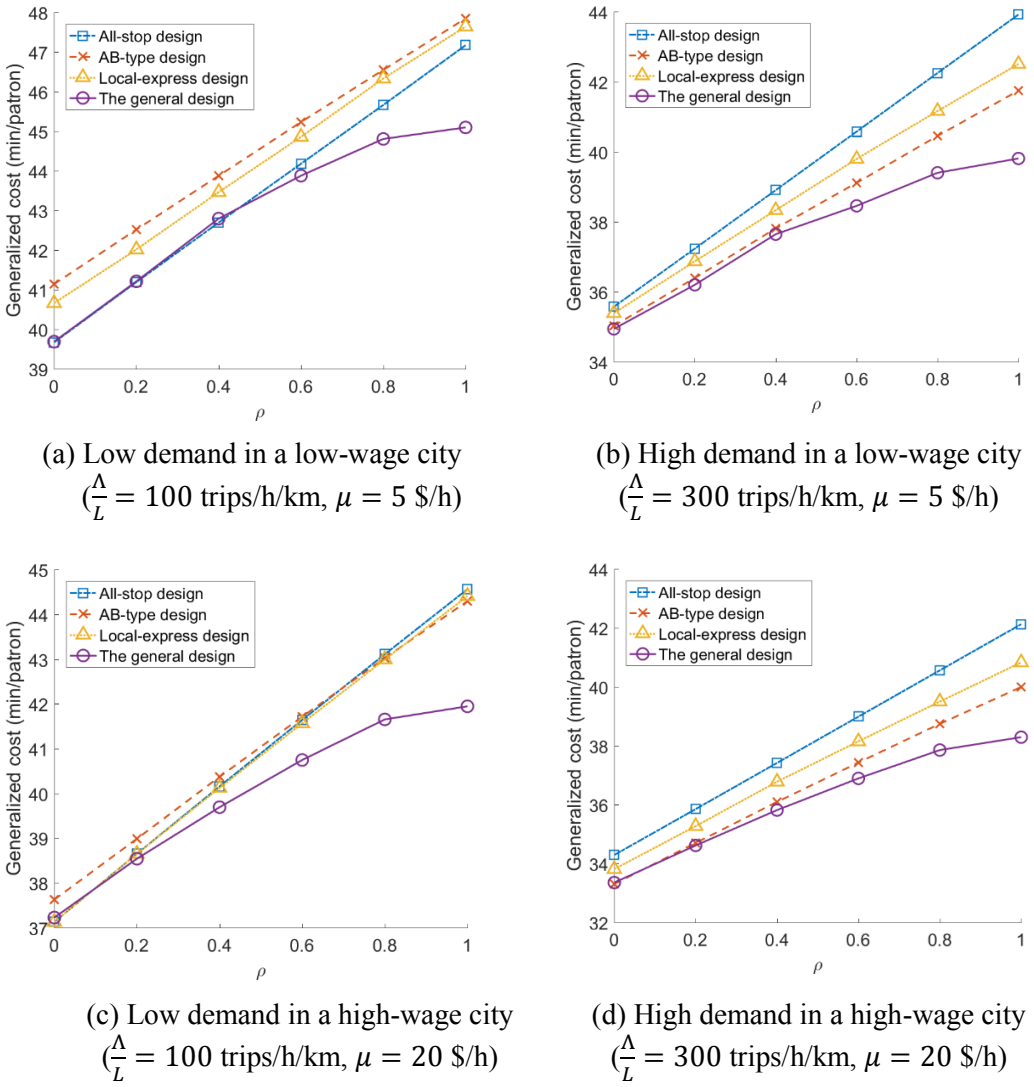
4.3.3 Parametric analysis

In this section, parametric analysis is conducted for general skip-stop service design of bus and rail systems, two time values (i.e. $\mu = 5$ \$/h and $\mu = 20$ \$/h), six degrees of demand heterogeneity (i.e. $\rho \in \{0,0.2,0.4,0.6,0.8,1\}$) and various average demand densities: $\frac{\Lambda}{L} = \{100, 200, 300\}$ trips/h/km for bus systems and $\frac{\Lambda}{L} = \{500, 700, 900\}$ trips/h/km for rail systems. The optimal general skip-stop service design is obtained for a 2-route system only, and is compared with three other forms of skip-stop service: AB-type service, local-express service, and all-stop service. Note for rail systems that the line infrastructure cost per km of local-express and general skip-stop services is twice of that of all-stop and AB-type services, since the former need a second track for train overtaking.

Figures 4.3a and b present the optimal generalized cost curves of the four service types for a bus corridor in a low-wage city ($\mu = 5$ \$/h) under low and high demand densities ($\frac{\Lambda}{L} = 100$ and 300 trips/h/km), respectively. The cost curves are plotted against degree of heterogeneity ($\rho \in [0,1]$). Figure 4.3a shows that under low demand the AB-type and local-express designs' performance is worse than the all-stop design (note that the AB-type designs examined in these figures consist of 2 or more routes in each direction). The figure also shows that the general design performs as good as the all-stop design for $\rho \leq 0.4$, and it outperforms all the three special design forms with greater cost savings as demand heterogeneity rises. Under high demand (Figure 4.3b), the all-stop design performs the worst, followed by the local-express design and the AB-type design. The general design still has the lowest cost regardless of ρ, and its cost saving increases with ρ.

To examine the effect of value of time, cost curves of the four service types are also plotted in Figures 4.3c-d for a high-wage city ($\mu = 20$ \$/h). Under low

demand, comparison between Figures 4.3a and c reveals that the skip-stop services are more favourable for high-wage cities; see in Figure 4.3c that the local-express and AB-type services now have similar costs as the all-stop service. Under high demand, however, the cost savings by skip-stop services are insensitive to the time value.



(a) Low demand in a low-wage city ($\frac{\Lambda}{L}$ = 100 trips/h/km, μ = 5 \$/h)

(b) High demand in a low-wage city ($\frac{\Lambda}{L}$ = 300 trips/h/km, μ = 5 \$/h)

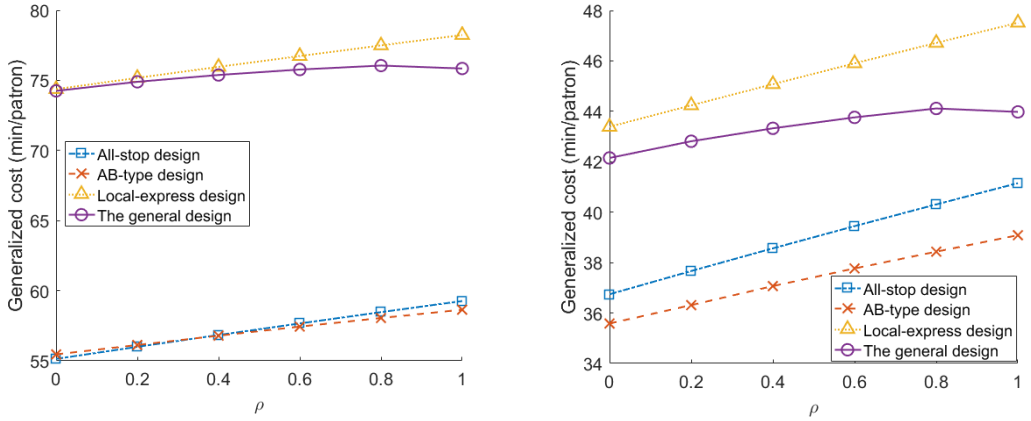
(c) Low demand in a high-wage city ($\frac{\Lambda}{L}$ = 100 trips/h/km, μ = 20 \$/h)

(d) High demand in a high-wage city ($\frac{\Lambda}{L}$ = 300 trips/h/km, μ = 20 \$/h)

Figure 4.3 Performance of different skip-stop service types for bus systems

Different findings are obtained for rail systems; see Figures 4.4a and b for the four service types' cost curves for a rail corridor under high demand ($\frac{\Lambda}{L}$ = 900 trips/h/km) in low and high-wage cities, respectively. Note now that the all-stop and AB-type services furnish much lower costs than local-express and general skip-stop services due to the additional infrastructure cost entailed by the latter two designs. The lowest-cost design is usually AB-type service. Comparison

between Figures 4.4a and b shows that the skip-stop services are more favourable for high-wage cities. Similar results are also observed for other demand patterns. They are however omitted here in the interest of brevity.



(a) High demand in a low-wage city ($\frac{\Lambda}{L} = 900$ trips/h/km, $\mu = 5$ \$/h)

(b) High demand in a high-wage city ($\frac{\Lambda}{L} = 900$ trips/h/km, $\mu = 20$\$/h)

Figure 4.4 Performance of different skip-stop service types for rail systems

4.3.4 Real-world case study

In this section, the four types of services are designed for a case study of Metro Line 1 in Chongqing City (China), as shown in Figure 4.5. The 35.55 km line contains 21 stops and has only one track in each direction. The stop-to-stop OD demand was collected during a peak hour (7:30-8:30) period on 2016.11.03. The two numbers in parentheses on a stop represent the stop's number (numbered from east to west) and longitudinal coordinate (coordinate of station 1 is set to zero), respectively. Figure 4.6 plots the boarding and alighting demand over the corridor in eastbound and westbound directions. Note that the boarding and alighting densities between stop 17 and stop 20 are significantly smaller than at other stops. The time value is set to $\mu = 20$ \$/h.

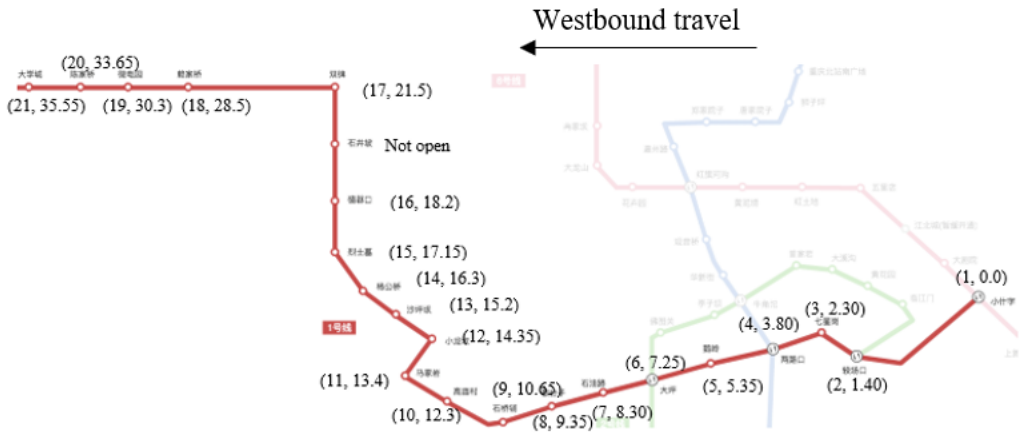


Figure 4.5 Layout of the Metro Line 1 of Chongqing

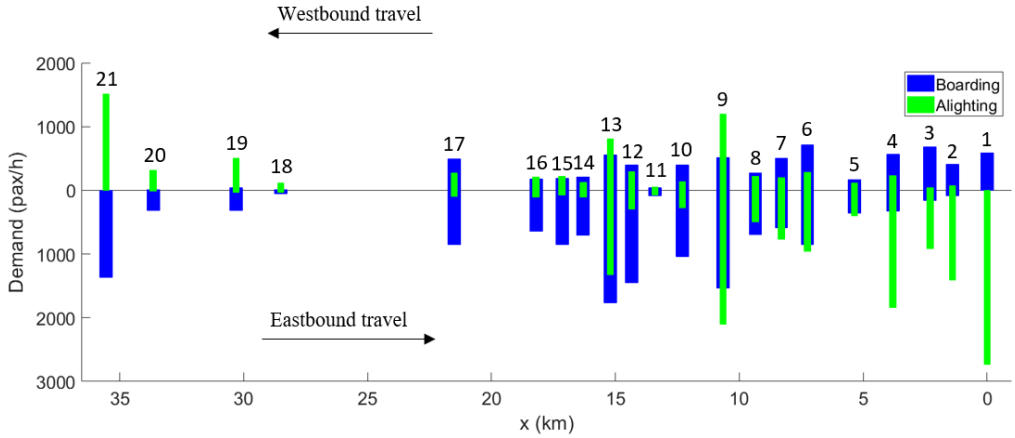


Figure 4.6 Boarding and alighting demand

Under the given demand pattern, the optimal general skip-stop design is shown in Figure 4.7, where the blue circles, red squares and black dots represent the A-, B- and T-type stops, respectively. The stop spacings vary in the range of $[0.4, 2.2]$ km. It is observed that the stop spacings between 20 km and 33 km are much larger than other parts of the corridor, and the skip-stop bay between 17 km and 26 km contains A-type stops only. These design features result from the significantly low demand between 20 and 34 km as shown in Figure 4.6.

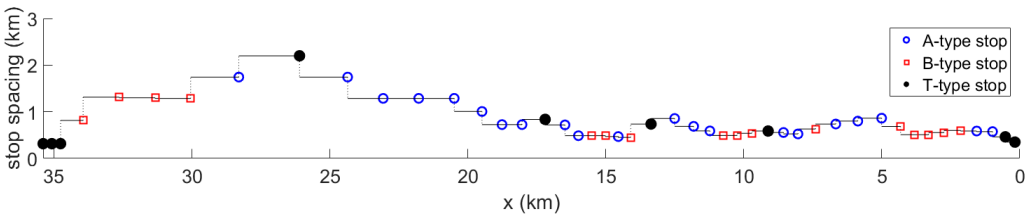


Figure 4.7 Optimal stop spacing of the general skip-stop design

Table 4.4 compares the optimal design results for the four service types and the existing design (the service headway of the existing design is optimized to provide a conservative comparison). Consistent with the findings in Section 4.3.3, the AB-type design furnishes the largest cost saving (29.3%) as compared to the existing system, since it can operate in the present single-track infrastructure. However, the general skip-stop design has the lowest user cost. This type of design would be the lowest-cost option if a double-track system is already built, or if the demand further increases to entail a double-track rail system.

Table 4.4 Comparison between four design types and the existing design

	Existing design	All-stop	AB-type	Local-express[#]	General skip-stop
Number of stops	21	40	48	39	46
H_1 (min)	4.2	3.2	4.5	7.5	4
H_2 (min)	-	-	4.5	4.4	3.7
User cost (min/patron)	48.5	30.6	28	29.4	27.1
Agency cost (min/patron)	12.6	14.5	15.2	24.6	25.8
Generalized cost (min/patron)	61.1	45.1	43.2	54	52.9
Cost saving as compared to the existing design	-	26.2%	29.3%	11.6%	13.4%

[#] H_1 corresponds to the headway of the local line

4.4 Summary of the General Skip-stop Design

A novel formulation has been developed to model a more general skip-stop design, combining the properties of discrete and CA models. This model allows both the transfer stops and non-transfer stops of different routes to be distributed along the corridor in an arbitrary fashion. The model takes a discrete OD matrix as input, and uses stop densities instead of individual stop locations as decision variables.

A heuristic method is proposed to solve the problem. A relaxed problem is also formulated to offer an initial solution to the heuristic method and a lower bound to the original problem. A recipe to convert the solution to a design with exact stop locations is provided too. The numerical tests verify that our heuristic method rapidly produces near-optimal solutions.

The superiority of this general design over all-stop, AB-type and local-express designs is demonstrated through many case studies (if the additional infrastructure cost for a second track or bus lane is ignored). Cost savings accrue up to 9.1% as compared against all-stop service, 7.2% against local-express service and 6.3% against AB-type service. These results suggest great application potential for our general skip-stop design model. However, for rail systems with expensive track cost, the AB-type design may still be the lowest-cost option if the demand is not too high, as is manifested by a real-world case study.

Chapter 5 Conclusions and Future Work

Section 5.1 summarizes this dissertation's contributions. Section 5.2 discusses possible extensions of the current work.

5.1 Major Contributions

Continuous approximation (CA) models for optimizing the design of two special skip-stop service forms, AB-type service and local-express service, were formulated. A novel discrete formulation for a more general skip-stop design was also developed. In all three of these models, stop locations and routing plan are jointly optimized under spatially heterogeneous demand. We believe these models mark an important advance of the present research frontier in skip-stop service optimization, since past studies either assumed unrealistic uniform demand pattern, or optimized skip-stop routing plan only for given stop locations, and they usually furnished heuristic solutions whose optimality gaps were difficult to evaluate. The thesis work also addressed other issues, including the modelling of both loop and linear corridors, asymmetric designs between two travel directions of a corridor, and patrons' route choice behaviors.

The two CA models were solved via applying calculus of variations to partially decompose the formulations by spatial coordinates. Parts of the formulations that cannot be locally decomposed (e.g. the backtracking-related cost items and the procedure for calculating patrons' route choice equilibrium) were updated via iterative steps. The discrete formulation for general design was solved by an iterative heuristic algorithm that employs an initial solution developed from a relaxed formulation. All these solution methods were demonstrated to be computationally efficient. Lower bounds were also developed for the AB-type and general design models, which were used to verify the near-optimality of model solutions.

Recipes were proposed to convert the solutions of stop spacings/densities to exact stop locations and routing plans, paving the way for real world implementation. These stop locations can be further refined under the consideration of local physical constraints, e.g. bridges and tunnels, ramps, junctions, and curbside parking space.

The advantages of the three skip-stop designs were examined through a large array of numerical experiments. Results show that skip-stop services significantly outperform conventional all-stop services under various demand patterns and operating conditions. In general, the cost savings resulting from optimally-designed skip-stop services increase with trip length, demand density, demand heterogeneity and time value.

Comparison between the three skip-stop service types reveals that: i) for single-mode corridors, the general skip-stop model produces the most flexible designs with the lowest generalized cost (if the additional infrastructure cost of a second track or bus lane is not included), but it also has the highest approximation error resulting from its specific formulation; ii) the local-express design is best operated with differentiated transit modes, which has substantially lower costs than single-mode systems; and iii) the AB-type service often furnishes the lowest-cost design for rail systems since it does not require an expensive second track. The advantages of above designs, especially the AB-type design, can be further enhanced by coordinating the schedules of different routes at transfer stops.

Admittedly, some of the above findings were obtained under limited demand patterns. There are infinite number of spatially heterogeneous demand patterns. More demand patterns (including real corridor cases) can be examined in the future, which may unveil new findings and more comprehensive understandings on the optimal structure of transit corridor design. Our models can also be easily tuned to optimize skip-stop routing plans for existing transit systems with fixed stop locations. Corridors with time-varying demand can be optimized by allowing for different service schemes during different periods, e.g. operating skip-stop service during peak hours and all-stop service during off-peak periods. These time-varying schemes may share the same set of stops.

5.2 Future Work

The following extensions of this thesis work can be conducted in the future:

i) More sophisticated, stochastic route choice models (e.g. logit-form models) can be incorporated into our modeling framework to furnish better

predictions of patrons' route choice behavior. Other stochastic factors in demand, patrons' wait time, and transit travel time will also be considered.

ii) The general skip-stop design model can be further tailored to include zonal service and short-turn service into the feasible design forms that the model can produce. This generalization will further expand the applicable scope of this model.

iii) The CA method proposed in this thesis can also be applied to model general, heterogeneous trunk-feeder corridors and networks consisting of a mix of various feeder service types (fixed-route buses, flex-route vans, shared bikes, etc.).

Appendix

A Tables of notation

Table A.1 List of common cost terms and parameters used in the paper

AUC_a	Patrons' total access and egress time for the all-stop service (h/h)
AUC_w	Patrons' total wait time for the all-stop service (h/h)
AUC_i	Patrons' total in-vehicle travel time for the all-stop service (h/h)
AAC_K	Distance-based vehicle operating cost for the all-stop service (h/h)
AAC_H	Time-based vehicle operating cost for the all-stop service (h/h)
AAC_I	Line infrastructure cost for the all-stop service (h/h)
AAC_S	Stop infrastructure cost for the all-stop service (h/h)
AC	Generalized cost for the all-stop service (h/h)
UC_a	Patrons' total access and egress time for the skip-stop service (h/h)
UC_w	Patrons' total wait time for the skip-stop service (h/h)
UC_i	Patrons' total in-vehicle travel time for the skip-stop service (h/h)
UC_t	Total transfer penalty for the skip-stop service (h/h)
AC_K	Distance-based vehicle operating cost for the skip-stop service (h/h)
AC_H	Time-based vehicle operating cost for the skip-stop service (h/h)
AC_I	Line infrastructure cost for the skip-stop service (h/h)
AC_S	Stop infrastructure cost for the skip-stop service (h/h)
SC	Generalized cost for the skip-stop service (h/h)
L	Corridor length (km)
v_w	Walking speed (km/h)
v	Vehicle cruise speed (km/h)
μ	Patrons' value of time ($/h)
H_{min}	Minimum vehicle headway (h)
H_{max}	Maximum vehicle headway (h)
K	Vehicle's passenger-carrying capacity (patron/veh)
C_t	Unit penalty cost per transfer (h)
π_v	Unit distance-based operating cost per vehicle-km ($/veh/km)
π_m	Unit time-based operating cost per vehicle-hour ($/veh/h)
π_i	Amortized construction and maintenance cost per km of line infrastructure (in one direction) per hour of service ($/km/h)
π_s	Amortized construction and maintenance cost per stop per hour of service ($/stop/h)

Table A.2 List of variables, functions, and parameters used in Chapter 2

Notation	Description
Decision variables/functions	
r_c	Number of skip-stop routes in the clockwise direction
r_{cc}	Number of skip-stop routes in the counterclockwise direction
$s(x)$	Stop spacing at x (km)
$k_c(x)$	The number of non-transfer stops experienced by each clockwise vehicle in the skip-stop bay containing x
$k_{cc}(x)$	The number of non-transfer stops experienced by each counterclockwise vehicle in the skip-stop bay containing x
H_c	Vehicle headway in the clockwise direction (h)
H_{cc}	Vehicle headway in the counterclockwise direction (h)
Demand variables/functions	
$\lambda(x, y)$	Demand density from origin x to destination y (trip/km^2/h)
Λ_c	Total demand of clockwise trips (trip/h)
Λ_{cc}	Total demand of counterclockwise trips (trip/h)
$P_c(x)$	Trip origin density of clockwise trips at x (trip/km/h)
$P_{cc}(x)$	Trip origin density of counterclockwise trips at x (trip/km/h)
$Q_c(x)$	Trip destination density of clockwise trips at x (trip/km/h)
$Q_{cc}(x)$	Trip destination density of counterclockwise trips at x (trip/km/h)
$o_c(x)$	On-board passenger flow of clockwise trips at x (trip/h)
$o_{cc}(x)$	On-board passenger flow of counterclockwise trips at x (trip/h)
l	Trip length (km)
$p_c(x)$	Probability density function of trip origins for clockwise trips
$p_{cc}(x)$	Probability density function of trip origins for counterclockwise trips
$\theta_c(l)$	Probability density function of trip length for clockwise trips
$\theta_{cc}(l)$	Probability density function of trip length for counterclockwise trips
σ_o	Standard deviation of the distribution of trip origins (km)
E_l	Mean trip length (km)
σ_l	Standard deviation of trip length (km)
Other parameters and variables	
$U_-^L(x, \delta)$	The left δ-neighborhood of x in a corridor of length L
$U_+^L(x, \delta)$	The right δ-neighborhood of x in a corridor of length L

$U^L(x,\delta)$	The δ-neighborhood of x in a corridor of length L, which is the union of both the left and right δ-neighborhoods of x
$T(x)$	Number of stops (including one transfer stop) in a skip-stop bay at x
$b_c(x)$	Density of backtracking trips at x in the clockwise direction (trip/km/h)
$b_{cc}(x)$	Density of backtracking trips at x in the counterclockwise direction (trip/km/h)
$\bar{\lambda}_c(x)$	Average density of clockwise trips contained in a skip-stop bay at x (trip/km/h)
$\bar{\lambda}_{cc}(x)$	Average density of counterclockwise trips contained in a skip-stop bay at x (trip/km/h)
τ	Vehicle dwell time at each stop (h)

Table A.3 List of variables, functions, and parameters used in Chapter 3

Notation	**Description**
Decision variables/functions	
$\delta_e(x)$	Stop density of express transit at location x
$\delta_l(x)$	Stop density of local transit at location x
H_e	Service headway of express transit
H_l	Service headway of local transit
Demand variables/functions	
$\lambda(x,y)$	Demand density from origin x to destination y (trip/km^2/h)
$\lambda_{(I)}^d(x,y)$	Demand density of route I in direction d from x to y (trip/km^2/h)
$P^d(x)$	Trip origin density in direction d at x (trip/km/h)
$Q^d(x)$	Trip destination density in direction d at x (trip/km/h)
$P_{(I)}^d(x)$	Trip origin density of route I in direction d at x (trip/km/h)
$Q_{(I)}^d(x)$	Trip destination density of route I in direction d at x (trip/km/h)
$B_l^d(x)$	Boarding density of local transits in direction d at x (trip/km/h)
$A_l^d(x)$	Alighting density of local transits in direction d at x (trip/km/h)
$B_e^d(x)$	Boarding density of express transits in direction d at x (trip/km/h)
$A_e^d(x)$	Alighting density of express transits in direction d at x (trip/km/h)
$\lambda_{e\to l}^d(x)$	Transfer density from express to local lines in direction d at x (trip/km/h)
$\lambda_{l\to e}^d(x)$	Transfer density from local to express lines in direction d at x (trip/km/h)
$\lambda_t^d(x)$	Transfer density in direction d at x (trip/km/h)

$o_e^d(x)$	On-board patron flows on express lines in direction d at x (trip/km/h)
$o_l^d(x)$	On-board patron flows on local lines in direction d at x (trip/km/h)
$o^d(x)$	On-board patron flows in direction d at x (trip/km/h)
Other parameters and variables	
I	Index of route types, $I \in \{l, e, le, el, lel\}$
d	Direction of trips, $d \in \{E, W\}$
$t_{(I)}^d(x, y)$	Travel time of route I in direction d from origin x to destination y (h)
$\kappa(x, y)$	Access and egress time from origin x to destination y (h)
$V_e^d(x)$	Commercial speeds of express line in direction d at x (km/h)
$V_l^d(x)$	Commercial speeds of local line in direction d at x (km/h)
$C_{(I)}^d(x)$	Critical distance of Route I in direction d at x (km)
$p_e(x)$	The probability that location x is closer to an express stop
$p_l(x)$	The probability that location x is closer to a local stop
τ_l^0	Constant delay at stops for local vehicles (h/stop)
τ_e^0	Constant delay at stops for express vehicles (h/stop)
τ_l^a	Alighting delay at stop for local vehicles (h/patron)
τ_e^a	Alighting delay at stop for express vehicles (h/patron)
τ_l^b	Boarding delay at stop for local vehicles (h/patron)
τ_e^b	Boarding delay at stop for express vehicles (h/patron)

Table A.4 List of variables, functions, and parameters used in Chapter 4

Notation	**Description**
Decision variables/functions	
$\phi_{k,r}$	Proportion of route r stop in segment k
t_k	Proportion of transfer stop in segment k
n_k	Number of stops in segment k
H_r	Service headway of route r
Demand variables/functions	
$\lambda_{i,j}$	Demand from segment i to j (trip/h)
$\bar{o}_{k,r}^d$	Average on-board patron flows of route r in segment k for direction d (trip/km/h)
$o_{k,r}^d$	On-board patron flows of route r at the right boundary of segment k for direction d (trip/km/h)

$\boldsymbol{U_k^d}$	Supplementary OD matrix for segment k for direction d
$u_{i,j}^{k,d}$	Element of $\boldsymbol{U_k^d}$ at row i and column j
$p_{i,j,r}$	Probability of a trip completed via route r from segment i to j
$\hat{p}_{i,j}$	Probability of a trip involving a transfer
b_k	Average backtracking density in segment k (trip/km/h)
B_k	Maximum backtracking flow in segment k in one direction (trip/h)
Other parameters and variables	
k	Index of segment
m	The number of segments
l_k	Length of segment k (km)
r	Index of route
R	Number of routes
d	Direction of trips, $d \in \{E, W\}$
$w_{i,j}$	Total wait cost for all the patrons from segment i to j (h/h)
s_k	Skip-stop bay length in segment k
s_{min}	Minimum stop length (km)
bl_k	Index of the left-most segment spanned by s_k
br_k	Index of the right-most segment spanned by s_k
Γ_k	Cost of trips without backtracking for segment k (h/h)
Θ_k	Additional cost of backtracking trips for segment k (h/h)

B Derivation of some patron cost components for the AB-type service

Patron wait time, in-vehicle travel time, and transfer penalty are derived in sections B.1, B.2, and B.3, respectively. For simplicity, only the derivation of these cost components associated with clockwise trips is presented. Counterclockwise trip costs can be derived similarly.

B.1 Patrons' wait time in equations (2.13-14b)

Skip-stop system trips can be classified into four types with respect to origin and destination stop types. The probabilities and average wait times for the four trip types are summarized as follows:

Type 1: Both the origin and destination stops are transfer stops. Among all the trips from x to y, the probability of Type 1 trips is $P_1(x,y) = \frac{1}{T(x)} \cdot \frac{1}{T(y)}$. These trips' average wait time is $W_1 = \frac{H_c}{2}$.

Type 2: The origin stop is a transfer stop and the destination stop is a non-transfer stop. The local probability of Type 2 trips is $P_2(x,y) = \frac{1}{T(x)} \cdot \frac{T(y)-1}{T(y)}$, and the average wait time is $W_2 = \frac{r_c H_c}{2}$.

Type 3: The origin stop is a non-transfer stop and the destination stop is a transfer stop. The local probability of Type 3 trips is $P_3(x,y) = \frac{T(x)-1}{T(x)} \cdot \frac{1}{T(y)}$, and the average wait time is $W_3 = \frac{r_c H_c}{2}$.

Type 4: Both the origin and destination stops are non-transfer stops. The local probability of Type 4 trips is $P_4(x,y) = \frac{T(x)-1}{T(x)} \cdot \frac{T(y)-1}{T(y)}$. Trips of this type can be further divided into three subtypes:

Type 4.1: The origin and destination stops are on the same route. The probability for Type 4.1 trips is $P_{41}(x,y) = \frac{1}{r_c} \frac{T(x)-1}{T(x)} \cdot \frac{T(y)-1}{T(y)}$, and the average wait time is $W_{41} = \frac{r_c H_c}{2}$.

Type 4.2: The origin and destination stops are on different routes but are located in the same skip-stop bay. Each Type 4.2 trip has a backtracking segment via a route in the opposite direction. The average wait time is thus given by $W_{42} = \frac{r_c H_c + r_{cc} H_{cc}}{2}$. Note that the local probability of these trips involving backtracking (similar to $P_{41}(x,y)$) cannot be simply written as a function of locations x and y. Thus, the approximate local density of these trips is calculated. First, $\bar{\lambda}_c(x)$ (trips/km/h) is denoted as the average density of clockwise "contained" trips at x. A contained trip is a trip whose origin and destination are both in the same skip-stop bay. The $\bar{\lambda}_c(x)$ is calculated by dividing the total number of contained trips in the skip-stop bay that contains x by that skip-stop bay's length, i.e.,

$$\bar{\lambda}_c(x) = \frac{\int_{z=U(x)}^{D(x)} \int_{y=z}^{D(x)} \lambda(z,y) dz dy}{D(x)-U(x)} \approx \frac{\int_{z \in U^L\left(x, \frac{T(x)s(x)}{2}\right)} \int_{y \in U^L\left(x, \frac{T(x)s(x)}{2}\right) \cap U_+^L(z, T(x)s(x))} \lambda(z,y) dz dy}{T(x)s(x)} \quad \text{(B1)}$$

where $U(x)$ and $D(x)$ denote the locations of the upstream and downstream transfer stops that enclose the skip-stop bay containing x. Note that $D(x) - U(x) = T(x)s(x)$. The approximation used in (B1) is obtained from $U(x) \approx x - \frac{T(x)s(x)}{2}$ and $D(x) \approx x + \frac{T(x)s(x)}{2}$.

Among those contained trips, the local probability that a trip involves backtracking is $\frac{r_c k_c(x)}{T(x)} \cdot \frac{(r_c-1)k_c(x)}{T(x)}$. The first fraction in the above formula is the probability that a trip's origin stop is a non-transfer stop in the bay; note that although a skip-stop bay has two transfer stops, for each of them only half of the catchment zone is contained in the bay. The second fraction is the probability that the trip's destination stop is a non-transfer stop in a bay on a different route from the origin stop. Note for a contained trip from x to y that $k_c(y) \approx k_c(x)$ since x and y are very close. Hence, the local density of backtracking trips is given by:

$$b_c(x) = \frac{r_c(r_c-1)k_c^2(x)}{T^2(x)} \bar{\lambda}_c(x) =$$

$$\frac{r_c(r_c-1)k_c^2(x)}{T^3(x)s(x)} \int_{z \in U^L\left(x, \frac{T(x)s(x)}{2}\right)} \int_{y \in U^L\left(x, \frac{T(x)s(x)}{2}\right) \cap U_+^L(z, T(x)s(x))} \lambda(z,y) dz dy$$

which is exactly equation (2.14a). The total number of backtracking trips is approximately $N_{42} = \int_{x=0}^{L} b_c(x) dx$.

Type 4.3: The origin and destination stops are located on different routes and in different skip-stop bays. The average wait time is $W_{43} = r_c H_c$ (including the transfer stop wait time). The number of Type 4.3 trips is the total number of Type 4 trips minus the number of Types 4.1 and 4.2 trips (for the same reason as above, a localized probability similar to $P_{41}(x,y)$ cannot be derived):

$$N_{43} = \iint_{D_c} \lambda(x,y) \left(1 - \frac{1}{r_c}\right) \frac{T(x)-1}{T(x)} \cdot \frac{T(y)-1}{T(y)} dx dy - \int_{x=0}^{L} b_c(x) dx,$$

where D_c denotes the shaded areas in Figure 2.1b, which represent the set of OD pairs (x,y) for clockwise trips; i.e. $D_c \equiv \left\{(x,y) | y \in U_+^L\left(x, \frac{L}{2}\right), x \in (0,L]\right\} = \left\{(x,y) | x \in U_-^L\left(y, \frac{L}{2}\right), y \in (0,L]\right\}$.

The total patrons' wait time for clockwise trips is therefore given by:

$$SUC_{w_c} = \iint_{D_c}(P_1(x,y)W_1 + P_2(x,y)W_2 + P_3(x,y)W_3 +$$

$$P_{41}(x,y)W_{41})\lambda(x,y)dxdy + N_{42}W_{42} + N_{43}W_{43} =$$

$$\iint_{D_c}\frac{H_c+r_cH_c(T(x)+T(y)-2)+(2r_c-1)H_c(T(x)-1)(T(y)-1)}{2T(x)T(y)}\lambda(x,y)dxdy +$$

$$\frac{r_{cc}H_{cc}-r_cH_c}{2}\int_{x=0}^{L}b_c(x)dx.$$

This can be simplified as follows:

$$UC_{w_c} = \frac{H_c}{2}\iint_{D_c}\lambda(x,y)\frac{1+r_cT(x)+r_cT(y)-2r_c+(2r_c-1)(T(x)T(y)-T(x)-T(y)+1)}{T(x)T(y)}dxdy +$$

$$\frac{r_{cc}H_{cc}-r_cH_c}{2}\int_{x=0}^{L}b_c(x)dx$$

$$= \frac{H_c}{2}\iint_{D_c}\lambda(x,y)\left(2r_c - 1 - \frac{r_c-1}{T(x)} - \frac{r_c-1}{T(y)}\right)dxdy + \frac{r_{cc}H_{cc}-r_cH_c}{2}\int_{x=0}^{L}b_c(x)dx$$

$$= \frac{(2r_c-1)H_c}{2}\iint_{D_c}\lambda(x,y)dxdy - \frac{(r_c-1)H_c}{2}\left(\iint_{D_c}\frac{\lambda(x,y)}{T(x)}dydx + \iint_{D_c}\frac{\lambda(x,y)}{T(y)}dxdy\right) +$$

$$\frac{r_{cc}H_{cc}-r_cH_c}{2}\int_{x=0}^{L}b_c(x)dx$$

$$= \frac{(2r_c-1)H_c}{2}\cdot\Lambda_c - \frac{(r_c-1)H_c}{2}\left(\int_{x=0}^{L}\frac{dx}{T(x)}\left(\int_{y\in U_+^L\left(x,\frac{L}{2}\right)}\lambda(x,y)dy\right) +\right.$$

$$\left.\int_{y=0}^{L}\frac{dy}{T(y)}\int_{x\in U_-^L\left(y,\frac{L}{2}\right)}\lambda(x,y)dx\right) + \frac{r_{cc}H_{cc}-r_cH_c}{2}\int_{x=0}^{L}b_c(x)dx$$

$$= \frac{(2r_c-1)H_c\Lambda_c}{2} - \frac{(r_c-1)H_c}{2}\left(\int_{x=0}^{L}\frac{P_c(x)}{T(x)}dx + \int_{y=0}^{L}\frac{Q_c(y)}{T(y)}dy\right) +$$

$$\frac{r_{cc}H_{cc}-r_cH_c}{2}\int_{x=0}^{L}b_c(x)dx$$

$$= \frac{(2r_c-1)H_c\Lambda_c}{2} - \frac{(r_c-1)H_c}{2}\int_{x=0}^{L}\frac{P_c(x)+Q_c(x)}{T(x)}dx + \frac{r_{cc}H_{cc}-r_cH_c}{2}\int_{x=0}^{L}b_c(x)dx \qquad \text{(B2)}$$

In the above derivation, the fourth equality uses the definition of Λ_c, and the fifth equality uses the definitions of $P_c(x)$ and $Q_c(y)$.

Similarly, the total wait time for counterclockwise trips is:

$$\frac{(2r_{cc}-1)H_{cc}\Lambda_{cc}}{2} - \frac{(r_{cc}-1)H_{cc}}{2}\int_{x=0}^{L}\frac{P_{cc}(x)+Q_{cc}(x)}{T(x)}dx + \frac{r_cH_c-r_{cc}H_{cc}}{2}\int_{x=0}^{L}b_{cc}(x)dx \quad \text{(B3)}$$

Combining (B2) and (B3), we have equation (2.13).

B.2 In-vehicle travel time in equations (2.15-16b)

The total in-vehicle travel time for clockwise trips consists of two parts:

1) Direct travel time. This part of travel time is in total:

$$\int_{x=0}^{L} o_c(x)\left(\frac{1}{v}+\frac{\tau(k_c(x)+1)}{T(x)s(x)}\right)dx,$$

where $\frac{\tau(k_c(x)+1)}{T(x)s(x)}$ is the average time spent on dwelling at stops per unit travel distance (note that $T(x)s(x)$ is the length of the skip-stop bay at x.

2) Backtracking travel time. The result derived by Gu et al. (2016) is used for the average extra distance traveled by a trip involving backtracking, which is $\frac{1}{3}T(x)s(x)$ for those trips contained in a skip-stop bay at x. Although Gu et al.'s result was derived assuming uniformly distributed demand, it can be used as a good approximation here because the slow-varying demand within a skip-stop bay can be approximately considered uniform. Note that half this distance is traveled in the clockwise direction and the other half is in the counterclockwise direction. Therefore, the average extra travel time per backtracking trip is:

$$\frac{1}{2}\left(\frac{1}{3}T(x)s(x)\cdot\left[\left(\frac{1}{v}+\frac{\tau(k_c(x)+1)}{T(x)s(x)}\right)+\left(\frac{1}{v}+\frac{\tau(k_{cc}(x)+1)}{T(x)s(x)}\right)\right]\right)=\frac{1}{3v}T(x)s(x)+\frac{1}{6}\tau(k_c(x)+k_{cc}(x)+2),$$

and the total extra travel time due to backtracking is:

$$\int_{x=0}^{L}\left[\frac{1}{3v}T(x)s(x)+\frac{1}{6}\tau(k_c(x)+k_{cc}(x)+2)\right]b_c(x)dx,$$

where $b_c(x)$ is the local density of clockwise backtracking trips, which is derived in Appendix B.1.

The above two cost items are summarized in equations (2.15-2.16b) showing the clockwise part of UC_i. The counterclockwise part can be similarly derived.

B.3 Transfer penalty in equation (2.17)

Each trip of Types 4.2 and 4.3 (see Appendix B.1) involves a transfer. Other trips have no transfer. Hence the total transfer numbers for clockwise trips is:

$\iint_{D_c} \lambda(x,y)\left(1-\frac{1}{r_c}\right)\frac{T(x)-1}{T(x)}\cdot\frac{T(y)-1}{T(y)}dxdy,$

where D_c denotes the set of OD pairs (x, y) for clockwise trips (see its definition in Appendix B.1).

This can be further simplified as:

$$\iint_{D_c} \lambda(x,y)\frac{r_c-1}{r_c}\left(1-\frac{1}{T(x)}\right)\left(1-\frac{1}{T(y)}\right)dxdy$$

$$=\frac{r_c-1}{r_c}\iint_{D_c} \lambda(x,y)\left(1-\frac{1}{T(x)}-\frac{1}{T(y)}+\frac{1}{T(x)T(y)}\right)dxdy$$

$$=\frac{r_c-1}{r_c}\left(\iint_{D_c} \lambda(x,y)dxdy-\iint_{D_c}\frac{\lambda(x,y)}{T(x)}dydx-\iint_{D_c}\frac{\lambda(x,y)}{T(y)}dxdy+\right.$$

$$\left.\iint_{D_c}\frac{\lambda(x,y)}{T(x)T(y)}dxdy\right)$$

$$=\frac{r_c-1}{r_c}\left(\Lambda_c-\int_{x=0}^{L}\frac{P_c(x)+Q_c(x)}{T(x)}dx+\iint_{D_c}\frac{\lambda(x,y)}{T(x)T(y)}dxdy\right)$$

$$\approx\frac{r_c-1}{r_c}\left(\Lambda_c-\int_{x=0}^{L}\frac{P_c(x)+Q_c(x)}{T(x)}dx+\frac{1}{2}\iint_{D_c}\left(\frac{\lambda(x,y)}{T^2(x)}+\frac{\lambda(x,y)}{T^2(y)}\right)dxdy\right)$$

$$=\frac{r_c-1}{r_c}\left(\Lambda_c-\int_{x=0}^{L}\frac{P_c(x)+Q_c(x)}{T(x)}dx+\frac{1}{2}\int_{x=0}^{L}\frac{P_c(x)+Q_c(x)}{T^2(x)}dx\right)$$

$$=\frac{r_c-1}{r_c}\left(\Lambda_c-\int_{x=0}^{L}\frac{\left(P_c(x)+Q_c(x)\right)(2T(x)-1)}{2T^2(x)}dx\right)$$

To see why the third equality in the above derivation holds, please refer to the derivation of equation (B2) in Appendix B.1.

The approximation in the next step is employed to reduce the double integral term to a single integral term, so that this cost component can be decomposed locally by x. The local decomposition is needed to ensure the solution method presented in Section 2.2.1 works. This approximation is also conservative since it overestimates the cost. The next equality is derived in similar fashion to the third equality of the above derivation.

Hence, the clockwise part of the transfer penalty cost is $C_t\frac{r_c-1}{r_c}\left(\Lambda_c-\int_{x=0}^{L}\frac{\left(P_c(x)+Q_c(x)\right)(2T(x)-1)}{2T^2(x)}dx\right)$, and the counterclockwise part can be derived similarly. This explains equation (2.17).

C Derivation of the lower bound for AB-type service

To derive the lower bound, the mathematical program (2.22a-f) is modified by: i) removing some cost terms from the objective function and reducing the objective function; and ii) relaxing some constraints.

First the cost terms incurred by backtracking are removed, which are part of the right-hand-sides of (2.13) (i.e. the patrons' waiting cost UC_w) and (2.15) (i.e. the patrons' in-vehicle travel cost UC_i). These cost terms are summed to:

$$\int_{x=0}^{L}\left(\frac{r_{cc}H_{cc}-r_cH_c}{2}\left(b_c(x)-b_{cc}(x)\right)+\left(\frac{1}{3v}s(x)T(x)+\frac{1}{6}\tau(k_c(x)+k_{cc}(x)+2)\right)\left(b_c(x)+b_{cc}(x)\right)\right)dx.$$

If the demand pattern is symmetric, $\frac{r_{cc}H_{cc}-r_cH_c}{2}\left(b_c(x)-b_{cc}(x)\right)=0$ and thus the above cost is positive. Otherwise, $\frac{r_{cc}H_{cc}-r_cH_c}{2}\left(b_c(x)-b_{cc}(x)\right)$ can be negative but its absolute value should be very small since it is the product of the differences between pairs of close numbers. Thus, the above cost, to be removed from the objective function, is still positive in most real cases.

For the remaining cost terms, note that $\frac{k_c(x)+1}{T(x)}=\frac{k_c(x)+1}{r_ck_c(x)+1}=\frac{1}{r_c}+\frac{1-1/r_c}{r_ck_c(x)+1}=\frac{1}{r_c}+\frac{r_c-1}{r_cT(x)}$ and $\frac{k_{cc}(x)+1}{T(x)}=\frac{1}{r_{cc}}+\frac{r_{cc}-1}{r_{cc}T(x)}$. Hence, we replace the left-hand-sides of the above equalities by the corresponding right-hand-sides in (2.16a-b) for UC_i and in (2.19) for AC_H. Further note that $-\frac{2T(x)-1}{2T^2(x)}\geq-\frac{1}{T(x)}$, and thus we replace $-\frac{2T(x)-1}{2T^2(x)}$ by $-\frac{1}{T(x)}$ in (2.17) for UC_t to further reduce the objective function. After rearranging the terms, the modified objective function can be written as:

$$SC_{LB}=\theta+\int_{x=0}^{L}\left(f(x)+\frac{\beta_c(x)+\beta_{cc}(x)}{T(x)}\right)dx$$

where θ is a function of scalar decision variables H_c, H_{cc}, r_c and r_{cc} only; $f(x)$, $\beta_c(x)$ and $\beta_{cc}(x)$ are functions of $s(x)$ as well as the scalar variables. These are presented as:

$$\theta = \frac{(2r_c-1)\Lambda_c}{2}H_c + \frac{(2r_{cc}-1)\Lambda_{cc}}{2}H_{cc} + C_t\frac{r_c-1}{r_c}\Lambda_c + C_t\frac{r_{cc}-1}{r_{cc}}\Lambda_{cc} + \frac{\pi_v L}{\mu}\left(\frac{1}{H_c}+\frac{1}{H_{cc}}\right) + \frac{2\pi_i L}{\mu}$$

$$f(x) = \frac{s(x)}{4v_w}\left(P_c(x)+Q_c(x)+P_{cc}(x)+Q_{cc}(x)\right) + \left(o_c(x)+\frac{\pi_m}{\mu H_c}\right)\left(\frac{1}{v}+\frac{\tau}{r_c s(x)}\right) + \left(o_{cc}(x)+\frac{\pi_m}{\mu H_{cc}}\right)\left(\frac{1}{v}+\frac{\tau}{r_{cc}s(x)}\right) + \frac{\pi_s}{\mu}\frac{1}{s(x)}$$

$$\beta_c(x) = \frac{\left(o_c(x)+\frac{\pi_m}{\mu H_c}\right)\tau}{s(x)}\frac{r_c-1}{r_c} - \left(\frac{r_c-1}{r_c}C_t + \frac{(r_c-1)H_c}{2}\right)\left(P_c(x)+Q_c(x)\right)$$

$$\beta_{cc}(x) = \frac{\left(o_{cc}(x)+\frac{\pi_m}{\mu H_{cc}}\right)\tau}{s(x)}\frac{r_{cc}-1}{r_{cc}} - \left(\frac{r_{cc}-1}{r_{cc}}C_t + \frac{(r_{cc}-1)H_{cc}}{2}\right)\left(P_{cc}(x)+Q_{cc}(x)\right).$$

The constraints (2.22d-e) are further relaxed by ignoring onboard patron flow incurred by backtracking trips (i.e. $B(x)$). The lower bound problem is formulated as:

$$\min_{r_c, r_{cc}, H_c, H_{cc}, s(x), T(x)} SC_{LB} = \theta + \int_{x=0}^{L}\left(f(x) + \frac{\beta_c(x)+\beta_{cc}(x)}{T(x)}\right)dx$$

subject to:

$r_c, r_{cc} \in \{1,2,3,4\}$

$$H_{min} + \tau \le H_c \le \frac{K}{\max\limits_{0<x\le L}\{o_c(x)\}}$$

$$H_{min} + \tau \le H_{cc} \le \frac{K}{\max\limits_{0<x\le L}\{o_{cc}(x)\}}$$

$s(x) > 0$

$T(x) \ge 1$

The solution to the problem constructs a lower bound to the solution of the original program (2.22a-f) because the objective function is reduced, and the constraints are relaxed. The remaining work is to find the above program's optimal solution.

The lower bound problem is solved using a bi-level method. At the lower level, we fix the scalar variables r_c, r_{cc}, H_c and H_{cc}, and optimize the integrand $f(x) + \frac{\beta_c(x)+\beta_{cc}(x)}{T(x)}$ for each $x \in [0, L]$. To achieve this, note that if $s(x)$ is also fixed, AC_{LB} will be minimized at either $T(x) = 1$ (if $\beta_c(x) + \beta_{cc}(x)$ is negative)

or $T(x) = \infty$ (if $\beta_c(x) + \beta_{cc}(x)$ is positive). The former case, $T(x) = 1$, indicates that all the stops are transfer stops; while the latter case indicates that there is no transfer stop throughout the corridor. Therefore, the integrand $f(x) + \frac{\beta_c(x)+\beta_{cc}(x)}{T(x)}$ can be minimized for each x as:

$$\min\left\{\min_{s(x)} f(x), \min_{s(x)} f(x) + \beta_c(x) + \beta_{cc}(x)\right\}.$$

Note further that $f(x)$, $\beta_c(x)$ and $\beta_{cc}(x)$ are all convex with respect to $s(x)$. Hence the global optimal solution to the above minimization problem can be easily solved by a gradient search method.

At the upper level, the scalar variables r_c, r_{cc}, H_c and H_{cc} are optimized via exhaustive search. r_c and r_{cc} can be easily enumerated from $\{1,2,3,4\} \times \{1,2,3,4\}$, and H_c and H_{cc} each will be searched from H_{min} to 15 min with a very small step, i.e. 0.1 min.

In summary, the optimal solution to the problem can be discovered by the method described above, and the solution is a lower bound to the optimal solution of (2.22a-f).

D. Derivation of the upper bound of s_k in the general skip-stop model

Here, we prove that $\frac{\alpha \cdot l_k}{t_k n_k + \frac{1}{L}}$ is a tight upper bound of $s_k = \min\left(L, x_k + \frac{l_k}{2t_k n_k}\right) - \max\left(0, x_k - \frac{l_k}{2t_k n_k}\right)$, where $\alpha = \max_k \frac{l_k+1}{l_k}$.

Proof. It is to prove that $\frac{\alpha \cdot l_k}{t_k n_k + \frac{1}{L}} \geq s_k$. The value of $\frac{l_k}{2t_k n_k}$ could be discussed in three cases:

Case 1: $\frac{l_k}{2t_k n_k} \geq L$. In this case, we can easily get $s_k = L - 0 = L$. Since $\frac{l_k}{2t_k n_k} \geq L \Rightarrow l_k \geq 2t_k n_k L$, we can derive that $\frac{\alpha \cdot l_k}{t_k n_k + \frac{1}{L}} \geq \frac{\alpha \cdot l_k}{\frac{l_k}{2L} + \frac{1}{L}} = \frac{2\alpha \cdot l_k}{l_k+2} L \geq \frac{2l_k+2}{l_k+2} L \geq L$.

Case 2: $\frac{L}{2} \leq \frac{l_k}{2t_k n_k} \leq L$. According to the value of x_k, there are four situations:

1) If $x_k + \frac{l_k}{2t_k n_k} \le L$ and $x_k - \frac{l_k}{2t_k n_k} \le 0$, then $x_k \le L - \frac{l_k}{2t_k n_k}$, $s_k = x_k + \frac{l_k}{2t_k n_k}$. Since $x_k + \frac{l_k}{2t_k n_k} \le L$, we have $s_k \le L$. Recognize that $\frac{L}{2} \le \frac{l_k}{2t_k n_k} \le L$, we can derive that $\frac{\alpha \cdot l_k}{t_k n_k + \frac{1}{L}} \ge \frac{\alpha \cdot l_k}{\frac{l_k}{L} + \frac{1}{L}} = \frac{\alpha \cdot l_k}{l_k + 1} L \ge L \ge s_k$.

2) If $x_k + \frac{l_k}{2t_k n_k} \le L$ and $x_k - \frac{l_k}{2t_k n_k} \ge 0$, then $x_k = \emptyset$. Hence, this situation does not exist.

3) If $x_k + \frac{l_k}{2t_k n_k} \ge L$ and $x_k - \frac{l_k}{2t_k n_k} \le 0$, then $L - \frac{l_k}{2t_k n_k} \le x_k \le \frac{l_k}{2t_k n_k}$, $s_k = L$. It is similar to situation 1), such that we have $\frac{\alpha \cdot l_k}{t_k n_k + \frac{1}{L}} \ge L$. Hence, $\frac{\alpha \cdot l_k}{t_k n_k + \frac{1}{L}} \ge s_k$.

4) If $x_k + \frac{l_k}{2t_k n_k} \ge L$ and $x_k - \frac{l_k}{2t_k n_k} \ge 0$, then $x_k = \emptyset$. Again, this situation does not exist.

Case 3: If $\frac{l_k}{2t_k n_k} \le \frac{L}{2}$, then let $x_k' = L - x_k$ and substitute x_k in s_k. Then the derivation appears similar to Case 2.

E Proof of Proposition 1 for the general skip-stop model

Proposition 1. Assuming that there is at least a feasible solution to problem (4.21), the minimum of problem (4.21) must be a lower bound of problem (4.20).

Proof. Since the constraints of problem (4.21) are tighter than those of problem (4.20), there must exist a feasible solution to problem (4.20). It is assumed that the optimal solution to (4.20) is $(\boldsymbol{\phi}^*, \boldsymbol{t}^*, \boldsymbol{n}^*)$ and the optimal solution to (4.21) is $(\boldsymbol{\phi}^{**}, \boldsymbol{t}^{**}, \boldsymbol{n}^{**})$. Checking the First-Order conditions of (4.20) with respect to n_k (n_k has only boundary constraints), we can easily verify that the optimal value of n_k must satisfy $n_k^* = \sqrt{a_k} \left[\frac{\pi_S}{\mu} + \sum_{r=1}^{R} \left(\tau\left(\bar{o}_{k,r}^E + \bar{o}_{k,r}^W\right) + \frac{2\pi_M \tau}{\mu} H_r^{-1}\right)\left(\phi_{k,r} + t_k\right)\right]^{-\frac{1}{2}}$. Since $\phi_{k,r} + t_k \le 1$, we have $n_k^* \ge \sqrt{a_k} \left(\frac{\pi_S}{\mu} + \tau(\bar{o}_k^E + \bar{o}_k^W) + \frac{2\pi_M \tau}{\mu} H^{-1}\right)^{-\frac{1}{2}} = n_k^{min}$. Therefore, $(\boldsymbol{\phi}^*, \boldsymbol{t}^*, \boldsymbol{n}^*)$ is a feasible solution of problem (4.21). Since $SC2(\boldsymbol{\phi}^*, \boldsymbol{t}^*, \boldsymbol{n}^*) \le SC1(\boldsymbol{\phi}^*, \boldsymbol{t}^*, \boldsymbol{n}^*)$, the minimum of (4.21)

$SC2(\boldsymbol{\phi}^{**}, \boldsymbol{t}^{**}, \boldsymbol{n}^{**})$ must be a lower bound of (4.20), satisfying $SC2(\boldsymbol{\phi}^{**}, \boldsymbol{t}^{**}, \boldsymbol{n}^{**}) \leq SC2(\boldsymbol{\phi}^{*}, \boldsymbol{t}^{*}, \boldsymbol{n}^{*}) \leq SC1(\boldsymbol{\phi}^{*}, \boldsymbol{t}^{*}, \boldsymbol{n}^{*})$.

References

Abdelhafiez, E. A., Salama, M. R., and Shalaby, M. A., 2017. Minimizing passenger travel time in URT system adopting skip-stop strategy. Journal of rail transport planning & management, 7(4), 277-290.

Belcher, J., 2015. Changes to Transit Service in the MBTA District 1964–2015. www.transithistory.org/roster/MBTARouteHistory.pdf.

Chen, H., Gu, W., Cassidy, M.J., and Daganzo, C.F., 2015. Optimal transit service atop ring-radial and grid street networks: A continuum approximation design method and comparisons. Transportation Research Part B, 81, 755-774.

Chen, J., Liu, Z., Zhu, S., and Wang, W., 2015. Design of limited-stop bus service with capacity constraint and stochastic travel time. Transportation Research Part E, 83, 1-15.

Chen, X., Hellinga, B., Chang, C. and Fu, L., 2015. Optimization of headways with stop skipping control: a case study of bus rapid transit system. Journal of Advanced Transportation, 49(3), 385-401.

Chicago-L.org., 2010, A/B Skip-Stop Express Service, https://www.chicago-l.org/operations/lines/route_ops/A-B.html

Chiraphadhanakul, V. and Barnhart, C., 2013. Incremental bus service design: combining limited-stop and local bus services. Public Transport, 5(1), 53-78.

Cipriani, E., Gori, S., and Petrelli, M, 2012. Transit network design: A procedure and an application to a large urban area. Transportation Research Part C, 20(1), 3-14.

Cortés, C.E., Jara-Díaz, S. and Tirachini, A., 2011. Integrating short turning and deadheading in the optimization of transit services. Transportation Research Part A, 45(5), 419-434.

Daganzo, C.F., 2007. Fundamentals of Transportation and Traffic Operations. Emerald Group Publishing Limited, UK.

Daganzo, C.F., 2009. A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. Transportation Research Part B, 43(10), 913-921.

Daganzo, C.F., 2010. Structure of competitive transit networks. Transportation Research Part B, 44(4), 434-446.

Delle Site, P. and Filippi, F., 1998. Service optimization for bus corridors with short-turn strategies and variable vehicle size. Transportation Research Part A, 32(1), 19-38.

Estrada, M., Roca-Riu, M., Badia, H., Robusté, F., and Daganzo, C.F., 2011. Design and implementation of efficient transit networks: procedure, case study and validity test. Transportation Research Part A, 45, 935–950.

Freyss, M., Giesen, R., and Muñoz, J.C., 2013. Continuous approximation for skip-stop operation in rail transit. Transportation Research Part C, 36, 419-433.

Furth, P.G., 1986. Zonal route design for transit corridors. Transportation Science, 20(1), 1-12.

Furth, P. G., 1987. Short turning on transit routes. Transportation Research Record, 1108, 42-52.

Ghoneim, N. S. A., and Wirasinghe, S. C. 1986. Optimum zone structure during peak periods for existing urban rail lines. Transportation Research Part B, 20(1), 7-18.

Ghoneim, N. S. A., and Wirasinghe, S. C. 1987. Optimum zone configuration for planned urban commuter rail lines. Transportation science, 21(2), 106-114.

Gu, W., Amini, Z., and Cassidy, M.J., 2016. Exploring alternative service schemes for busy transit corridors. Transportation Research Part B, 93, 126-145.

Hidalgo, D., Lleras, G., and Hernández, E., 2013. Methodology for calculating passenger capacity in bus rapid transit systems: Application to the TransMilenio system in Bogotá, Colombia. Research in Transportation Economics, 39(1), 139-142.

Hirsch, L., Jordan, J., Hickey, R., and Cravo, V., 2000. Effects of fare incentives on New York City transit ridership. Transportation Research Record, (1735), 147-157.

Holryod, E., 1967. Optimum bus service: a theoretical model for a large uniform urban area. In: Edie, L. (Ed.), Proceedings of the Third International Symposium on the Theory of Traffic Flow. Elsevier, 308–328.

Hurdle, V.F. and Wirasinghe, S.C., 1980. Location of Rail Stations for Many-to-One Travel Demand and Several Feeder modes. Journal of Advanced Transportation, 14, 29-46.

Jiang, F., Cacchiani, V., and Toth, P., 2017. Train timetabling by skip-stop planning in highly congested lines. Transportation Research Part B, 104, 149-174.

Jordan, W.C. and Turnquist, M.A., 1979. Zone scheduling of bus routes to improve service reliability. Transportation Science, 13(3), 242-268.

Larrain, H., 2013. Diseño de Servicios Expresos para Buses. PhD diss., Pontificia Universidad Católica de Chile, Santiago, Chile.

Larrain, H., Muñoz, J.C. and Giesen, R., 2015. Generation and design heuristics for zonal express services. Transportation Research Part E, 79, 201-212.

Larrain, H., and Muñoz, J.C., 2016. When and where are limited-stop bus services justified? Transportmetrica A, 12(9), 811-831.

Lee, Y. J., Shariat, S., and Choi, K., 2014. Optimizing skip-stop rail transit stopping strategy using a genetic algorithm. Journal of Public Transportation, 17(2), 7.

Leiva, C., Muñoz, J.C., Giesen, R., and Larrain, H., 2010. Design of limited-stop services for an urban bus corridor with capacity constraints. Transportation Research Part B, 44(10), 1186-1201.

Li, H. and Bertini, R., 2009. Assessing a model for optimal bus stop spacing with high-resolution archived stop-level data. Transportation Research Record, 2111, 24-32.

Li, Y., Rousseau, J. M., and Gendreau, M., 1995. Real-time dispatching of public transit operations with and without bus location information. Computer-Aided Transit Scheduling. Springer, Berlin, Heidelberg, 296-308.

Liu, Z., Yan, Y., Qu, X., and Zhang, Y., 2013. Bus stop-skipping scheme with random travel time. Transportation Research Part C, 35, 46-56.

Medina, M., Giesen, R., Muñoz, J., 2013. Model for the optimal location of bus stops and its application to a public transport corridor in Santiago, Chile. Transportation Research Record, 2352, 84-93.

Mejias, L., and Deakin, E., 2005. Redevelopment and revitalization along urban arterials: Case study of San Pablo avenue, California, from the developers' perspective. Transportation Research Record, 1902, 26-34.

Meng, Q. and Qu, X., 2013. Bus dwell time estimation at bus bays: A probabilistic approach. Transportation Research Part C, 36, 61-71.

Metro de Santiago, 2008. Transport Supply Optimization.

Newell, G.F. and Potts, R.B., 1964. Maintaining a bus schedule. Proceedings of the 2nd conference of the Australian Road Research Board, 388-393.

Newell, G.F., 1971. Dispatching policies for a transportation route. Transportation Science, 5, 91–105.

Newell, G.F., 1973. Scheduling, location, transportation, and continuum mechanics: some simple approximations to optimization problems. SIAM Journal on Applied Mathematics, 25(3), 346-360.

Newell, G.F., 1979. Some issues relating to the optimal design of bus routes. Transportation Science, 13(1), 20–35.

Niu, H., 2011. Determination of the skip-stop scheduling for a congested transit line by bilevel genetic algorithm. International Journal of Computational Intelligence Systems 4(6), 1158-1167.

Qi, J., Yang, L., Di, Z., Li, S., Yang, K., and Gao, Y., 2018. Integrated optimization for train operation zone and stop plan with passenger distributions. Transportation Research Part E, 109, 151-173.

Sheffi, Y., 1985. Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. New Jersey: Prentice-Hall, Inc.

Sivakumaran, K., Li, Y., Cassidy, M. and Madanat, S., 2014. Access and the choice of transit technology. Transportation Research Part A, 59, 204-221.

Soto, G., Larrain, H., and Muñoz, J. C., 2017. A new solution framework for the limited-stop bus service design problem. Transportation Research Part B, 105, 67-85.

Thilakaratne, R.S. and Wirasinghe, S.C., 2016. Implementation of Bus Rapid Transit (BRT) on an optimal segment of a long regular bus route. International Journal of Urban Sciences, 20(1), 15-29.

Tirachini, A., Cortés, C.E. and Jara-Díaz, S.R., 2011. Optimal design and benefits of a short turning strategy for a bus corridor. Transportation, 38(1), 169-189.

Turnquist, M. A., 1979. Zone scheduling of urban bus routes. Journal of Transportation Engineering, 105(1), 1-13.

Ulusoy, Y., Chien, S. and Wei, C.H., 2010. Optimal all-stop, short-turn, and express transit services under heterogeneous demand. Transportation Research Record, 2197, 8-18.

Vaughan, R.J. and Cousins, E.A., 1977. Optimum location of stops on a bus route. Proceedings of 7th International Symposium of Transportation and Traffic Theory, 697-719.

Vaughan, R.J., 1986. Optimum polar networks for an urban bus system with a many-to-many travel demand. Transportation Research Part B, 20B(3), 215–224.

Vuchic, V.R., 2007. Urban Transit System and Technology. Hoboken, NJ: John Wiley & Sons.

Wirasinghe, S.C., Hurdle, V., Newell, G., 1977. Optimal parameters for a coordinated rail and bus transit system. Transportation Science, 11, 359–374.

Wirasinghe, S.C., 1980. Nearly optimal parameters for a rail/feeder bus system on a rectangular grid. Transportation Research Part A, 14(1), 33-40.

Wirasinghe, S.C. and Ghoneim, N.S., 1981. Spacing of bus-stops for many to many travel demand. Transportation Science, 15(3), 210-221.

Yang, L., Qi, J., Li, S., and Gao, Y., 2016. Collaborative optimization for train scheduling and train stop planning on high-speed railways. Omega, 64, 57-76.

Yue, Y., Wang, S., Zhou, L., Tong, L., and Saat, M. R., 2016. Optimizing train stopping patterns and schedules for high-speed passenger rail corridors. Transportation Research Part C, 63, 126-146.