Transportation Research Part B 128 (2019) 408–434

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Flexible parking reservation system and pricing: A continuum approximation approach

Xiaotian Wang[a], Xin Wang[b,*]

[a] Department of Industrial and Systems Engineering, University of Wisconsin-Madison, United States
[b] Department of Industrial and Systems Engineering & Grainger Institute for Engineering, University of Wisconsin-Madison, United States

A R T I C L E I N F O

Article history:
Received 20 December 2018
Revised 22 May 2019
Accepted 13 July 2019
Available online 29 August 2019

Keywords:
Continuum approximation
Parking reservation management
Stackelberg competition,

A B S T R A C T

Reservation-based parking systems have the merit of eliminating vehicle cruising for parking. While many long-period (e.g., daily) parking reservation services are already in use, short-period (e.g., hourly) parking reservation remains a huge challenge due to the high uncertainty of customer arrivals and departures. To mitigate the service failure caused by random late departures of customers, we propose a new flexible reservation mechanism in which the reservation is no longer restricted to a specific location at a specific time, but tolerates predetermined spatiotemporal flexibility instead. With a pricing instrument designed for such parking flexibility, customers can coordinate to significantly reduce the reservation failure rate, resulting in an optimal system equilibrium benefiting the entire society. Due to the complex nature of this system, a continuum approximation framework is used to provide tractable analysis for a large-scale urban parking system. We can successfully provide accurate system management decision support with a bounded optimality gap and analytical insights.

1. Introduction

Parking in urban areas has become a major problem in many large cities. According to the study by Shoup (2006), cruising for curb parking accounts for around 30% of traffic in congested urban areas and takes on average 8.1 min, wasting a significant amount of time and fuel. Additionally, the environmental impact is also substantial. For vast urban areas such as Chicago, cruising for parking causes over 129,000 tons of CO_2 emission each year (Ayala et al., 2012). These economic and environmental issues have raised many interests in designing a more efficient parking system.

The rapid development of information and communication technology has enabled various advances in parking management. First, dynamic pricing has been introduced to control the demand (Qian and Rajagopal, 2014). A concrete example of its implementation is the SFpark project in San Francisco. Second, parking reservation systems have become available, which significantly reduce the cruising time and improve the driver's experience. Such an approach further reduces urban congestion (Yang et al., 2013). As listed in Chen et al. (2015), there are a number of parking reservation applications in the market such as SpotHero, ParkWhiz, ParkMe, and Parking Panda, which provide an integrated platform for both pricing and reservation management. Such convenient services further allow the integration of dynamic pricing and parking reservation (Mackowski et al., 2015; Lei and Ouyang, 2017), with a great potential to improve parking efficiency.

* Corresponding author.
E-mail address: xin.wang@wisc.edu (X. Wang).

https://doi.org/10.1016/j.trb.2019.07.006

Table 1
Illustrative example of a parking reservation service with flexibility.

Scenarios	I	II	III	IV
Late departure time of A (minute)	2	2	4	4
Longest time C is willing to wait (minute)	0	3	3	3
Longest distance C is willing to detour (unit)	0	0	0	1
C's parking outcome	Failure	Spot 1	Failure	Spot 2

Many studies on parking management have been conducted with a specific focus on parking for a long period of time (e.g., the whole day-time). However, parking becomes particularly problematic when seeking short-period parking spots. Therefore, managing the parking system under the continuous arrivals and departures of short-period parking demand is a key factor in solving the urban hassle. Recent works of Mackowski et al. (2015) and Lei and Ouyang (2017) considered the short-period parking problem, but under the implicit assumption that users will abide by their reservations, e.g., departing before the end of their reservation periods. However, in the real world, the uncertainty in customers' parking behavior makes this assumption questionable. In 2017, New York City issued over 1.5 million tickets for parking in excess of the allowed time (NYC-OpenData, 2017). In case of the parking reservation, such late departing behavior could lead to the service failure of the parking reservation system, which causes unpleasant experiences in people's everyday life, especially when parking resources are in short supply. For example, let us imagine that a customer reserved a parking spot in the downtown area from 10:00 a.m. to 11:00 a.m. for a dentist appointment at 10:15 a.m. He managed to arrive at the reserved spot on time, but found out the spot was still occupied by the user ahead of his reservation. There might be a certain compensation policy addressing such cases, but at the moment he had to park quickly so he wouldn't miss the appointment. At this point he had a few options. He could wait for the spot to become available, hoping the current user would leave soon. Otherwise, he could cruise to find a vacant spot without any reservation, which would be a difficult task in the downtown area. Lastly, he could pay a much higher fee for parking in a nearby garage, if there was one close enough for walking back to catch the appointment. Obviously none of these cases would be a pleasant experience for him. Another problematic example would be electric cars that need to be charged at the reserved spot. For these drivers it would be a major inconvenience to arrive in time with a low battery level and find out the spot is still occupied. To avoid the service failure and the customer's subsequent unsatisfying experience, addressing the uncertainty inherent in customers' behavior is critical for a parking reservation system.

To mitigate the effects of parking uncertainty, we propose a new mechanism to control both supply and demand. On the supply side, we apply a strategy similar to *shared parking*. Shao et al. (2016) depicted shared parking as a strategy to improve parking space utilization by sharing parking spaces among different users, given the fact that most parking spaces are only used part-time. For example, during off-peak hours, a restaurant can open its parking spots to customers of a nearby movie theater. In the case of our parking reservation system, when parking spaces are not fully reserved in a neighborhood, the vacant spots could be seen as the shared spaces and be assigned to customers whose reserved spots are not available upon their arrivals. To be specific, the system might relocate users to nearby parking spaces to avoid the service failure. On the demand side, the system might ask users to wait for a short period of time for the potential releasing of occupied spaces. Under such a spatiotemporal flexible mechanism, the system service failure can be reduced to a certain threshold. To comprehend the benefit of the flexible mechanism, we construct a simple two-node example. Consider the case where two parking spots are one unit distance apart. For the first hour, Spots 1 and 2 are reserved by Customers A and B, respectively. For the second hour, Spot 1 is reserved by Customer C, while Spot 2 is empty. Customer A departs late, while customer B departs on time. Four scenarios are considered in Table 1, where Scenario I represents the traditional parking reservation system, where customer C will not be able to park due to the late departure of customer A, resulting in a service failure. Scenario II shows that if Customer C can tolerate a certain waiting time, e.g., 3 min, he/she can park successfully. In Scenario III, even if Customer C can wait a short time, there is still a chance that A departs extremely late, resulting in a service failure as well. While in scenario IV, if Customer C accepts to park at the nearby Spot 2, the parking can be successful despite A departing extremely late. Such a simple example shows a certain tolerance in waiting time and detour distance can greatly improve the parking success rate under uncertain late departures.

It is intuitive that the relocation distance and the waiting time have to be restricted to the user's acceptable level. To this end, in the remaining context of the paper we introduce two additional pricing tools of the system to affect a user's decision about the maximum relocation distance and the maximum waiting time, respectively named as *region flexibility* and *time flexibility*.

With region and time flexibility, we now seek to establish a reliable parking reservation system. To capture the user behavior on parking pricing, a bilevel Stackelberg leader-follower game is modeled. Due to the high nonconvexity and the curse of dimensionality of the problem, we develop a continuum approximation (CA) based framework to provide a feasible solution with a bounded optimality gap. Numerical tests show that CA solutions are near-optimal, and the proposed flexible parking reservation system significantly beats its non-flexible counterpart. Furthermore, a case study of a large urban area of San Francisco is conducted, showing the efficiency of applying the proposed model and algorithm to large-scale problems. Managerial insights are drawn afterwards.

We summarize our contributions as follows:

1. An innovative flexible parking reservation system is proposed to mitigate the service failure in traditional short-period parking reservation systems. Concepts of spatiotemporal flexibility are introduced to enlarge the degree of freedom in parking demand and supply management.
2. A bilevel Stackelberg game-theoretical model is established to capture the parking agency's decision process, incorporating the customer's behavior. A multi-stage stochastic process is further introduced to address the challenges of dynamics in short-term reservation with highly uncertain customer departures.
3. A neural network based continuum approximation method is developed to conquer the highly non-convex, large-scale problem structure, where the neural network reduces the complex multi-stage stochastic programming into a local analytical functional, and CA further provides feasible solutions to the overall problem. This method provides a general framework to balance the efficiency and accuracy for similar types of problems.
4. Fruitful insights are generated for parking management in reality through a case study of the San Francisco downtown area and its comparison to the existing SFpark project. The proposed flexible reservation scheme is extremely effective and results in significant benefit to society in most cases.

The rest of this paper is organized as follows. Section 2 provides a literature review on parking problems and solution methodologies. Section 3 explains the notation and the formulation of the proposed flexible parking reservation system. Section 4 presents the solution algorithm based on CA and the optimality gap obtained from perfect information relaxation. Section 5 illustrates the numerical results and provides some managerial insights. Section 6 concludes the paper and discusses possible future research.

2. Literature review

In the interest of our research, a large portion of parking management strategies can be divided into two main categories, parking information management and parking demand management. Research in parking information management utilizes the increasing amount of data provided by the developing information and communication technology, such as the sensors installed on parking spots, to improve the efficiency of the parking system. Parking guidance systems provide users with the real time availability information about the parking spaces and guide users to their destinations. Such a system has been implemented in many cities, e.g., Pittsburgh, Chicago, San Jose etc. (Idris et al., 2009). Based on the parking availability information, parking reservation further enables users to secure the available parking spaces before their arrivals. Such reservation strategies have been studied to reduce the traffic congestion. For example, Yang and Wang (2011) explored a system of tradeable travel credits to manage network mobility, where credits are initially distributed by the governments to all travelers, and later can be traded among travelers for the use of passing links in the urban network. In the spirit of their work, Zhang et al. (2011) studied the policies of parking permits allocation and free trade of parking permits for managing parking under limited parking spaces. Following the idea of parking permits, studies on parking reservation were conducted. Chen et al. (2015) proposed a parking reservation system with a simple reservation scheme to minimize the total social cost associated with parking. Yang et al. (2013) concluded that under the morning commute bottleneck model, an appropriate combination of reserved and unreserved parking spots can relieve traffic congestion and hence reduce the total system cost. Liu et al. (2014) considered a parking reservation scheme with expiration time, where commuters with a parking reservation have to arrive at parking spots before the predetermined expiration time.

The other category of parking advances is demand management. Most research focused on pricing strategies. Shoup (2005) illustrated the enormous impacts of parking pricing on urban transportation systems. Arnott et al. (1991) pointed out that location-dependent parking fees effectively increase the social welfare. Anderson and De Palma (2004) treated parking as a common property resource and examined the benefit of pricing it. Qian et al. (2012) studied how parking fee and parking supply can be optimized to alleviate traffic congestion and reduce total social costs. He et al. (2015) considered a parking competition problem and discussed the optimal pricing scheme to reach the optimal assignment of parking spaces. Recently, the role of dynamic pricing in parking management was studied by some researchers. Qian and Rajagopal (2014) proposed a dynamic pricing model to obtain the optimal parking pricing under demand uncertainty through a stochastic control problem, where they assumed real-time sensing of demand is available. Qian and Rajagopal (2015) further extended the study to the case when travelers make parking location choices and departure time choices. Zheng and Geroliminis (2016) proposed a macroscopic fundamental diagram based model to capture the dynamics at the network level, and developed a pricing strategy to reduce congestion. The integration of parking reservation and dynamic pricing emerged recently. Mackowski et al. (2015) proposed a dynamic parking pricing model for on- and off-street parking and concluded that their model reduced parking externalities significantly. Lei and Ouyang (2017) developed an integrated dynamic parking pricing and reservation mechanism and showed that it outperformed the myopic policy. However, based on our knowledge, all reservation-based parking management literature so far assumes customers will behave according to their reservations and depart on time, which will not hold in reality. Our work fills this research gap to address the critical service failure due to late departures, which is essential to establish a short-term parking reservation system.

There is also a small amount of literature that studied shared parking, which can be seen as a strategy to increase the supply of parking slots without expanding the physical capacity. Shao et al. (2016) designed a platform to allocate rentable private parking lots to drivers under the parking reservation environment. Xu et al. (2016) designed price-compatible matching mechanisms to address the shared parking problem. Xiao et al. (2018) proposed truthful double auction mechanisms for the parking sharing platform. In our work, the proposed flexible parking reservation mechanism uses price to control demand and supply at the same time, in order to achieve high parking resource's utilization and maintain a high level of service completion under uncertain customer parking behavior.

From the methodological point of view, our model follows the typical Stackelberg leader-follower game. Such game-theoretical structure has been widely used in modeling many customer behavioral problems. Examples include the optimization of road tools in Yang and Bell (1997), the transportation network design problem in Chiou (2005), the role of government policies in the biofuel supply chain design in Bai et al. (2016) and biofuel industry development in Wang et al. (2017). Due to the difficulty of the problem structure, CA based techniques are considered and research has shown its success in a wide range of fields. As described in Ansari et al. (2017), the key idea of CA is to continuously represent the discrete input data and decision variables, and approximate the objective into a functional (e.g., integration) of localized problems. Then the overall problem reduces to a set of point-wise homogeneous problems, each can be solved based on symmetrical parameter settings. By construction, the solution of CA is often in an analytical form and easy to reveal managerial insights. In addition, solving a CA model is generally much more efficient compared to its discrete counterpart. After first being proposed by Newell (1971), CA has been extensively studied and applied to various logistics problems, including facility location problems (Ouyang and Daganzo, 2006; Wang and Ouyang, 2013; Wang et al., 2016), distribution and transit problems (Smilowitz and Daganzo, 2007; Ouyang et al., 2014), and integrated supply chain and logistics studies (Bai et al., 2015; Lim et al., 2016). Readers are referred to Langevin et al. (1996) and Ansari et al. (2017) for a more complete review of CA studies. Finally, to evaluate the feasible solution obtained from CA, we calculate an optimality gap by assuming perfect information. This idea of information relaxation has been studied in stochastic programming literature such as Rockafellar and Wets (1991) and Brown et al. (2010).

3. Modeling

In traditional parking reservation systems without any flexibility, the realization of such systems is direct and easy. Let the customer park when the reserved parking lot is available; otherwise, a service failure occurs. On the contrary, to achieve the reservation flexibility, the system needs to incorporate both strategic decisions, such as how to set the reservation pricing scheme, and operational decisions, such as how to realize a final timing and location assignment for reserved customers once they arrive. To achieve this, we divide the flexible reservation system into the following two phases.

In the first phase, we focus on the reservation contract between our system and parking customers, which is relatively macroscopic. In particular, our system determines a parking price for each spot at each time, with certain discounts due to the maximum relocation distance and the maximum time delay in the final assignment. Then, all customers aiming to park near a specific location and time (e.g., in the aforementioned dentist example, all customers who need to visit the clinic at 10:00 a.m.) jointly react to such a pricing bundle. Similar to a dynamic parking pricing problem, the system can manage the parking demand through pricing adjustments. When the system raises the parking price, parking demand will reduce. Meanwhile, due to the introduction of parking flexibility, the system can also alter the discount for flexibility. If a significant discount is provided to customers per unit increase in maximum relocation radius or delay, customers may well take the discount at a sacrifice of their convenience. Hence, a general disutility should be used to evaluate the parking behavior, which is captured by a Customer Problem in Section 3.2. Note that we focus on the integrated parking customers' behavior, hence a single customer's action is considered as infinitesimal.

In the second phase, we focus on the realization of the reservation contract, which is microscopic. This phase aims to evaluate how the flexibility contract impacts the service failure of the final assignment. Therefore, we need to trace down to each customer's assignment. The joint customer demand equilibrium in the first phase can be used to sample the actual demand reservations. Then we build the Assignment Problem in Section 3.3 to optimize parking assignments given these reservations. The corresponding service failure measure will be used to help the first phase decisions. For the ease of reading, the notation for variables, parameters and functions used in the paper are summarized in Appendix A.

3.1. Problem settings

We consider a set of parking lots $\mathcal{J}$ spatially distributed over an urban area, providing potential parking services for the public over an infinite planning horizon $[0, +\infty)$. Without loss of generality, we assume the capacity of each parking lot is one, only allowing at most one vehicle to park at a specific time period (e.g., public parking meters). Suppose the parking privilege at each lot is packed into equal time intervals for sale, with a length l for each slot, i.e., a customer can choose to park at $[0, l], [l, 2l], \ldots$. In addition, we require that customers must book parking slots in advance for at least $(n-1)l$ long, $n \in \mathbb{Z}_+$, e.g., a customer can book a slot at $[(n-1)l, nl]$ or later for a lot $j \in \mathcal{J}$ at time 0. There is a centralized parking reservation system that handles all reservation requests. We consider a flexible assignment policy, where the system does not immediately secure the reserved parking slot for the customer but waits until the reserved time comes. Therefore, the system always makes assignment decisions with the reservation information in the following nl time periods (n slots). Our following analysis will focus on a selected particular rolling horizon, e.g., $\mathcal{T} = [0, nl]$, without loss of generality.

We now introduce a flexible parking mechanism. Suppose a customer reserved a parking lot j at $[(k-1)l, kl]$, $1 \le k \le n$. When reserving the parking lot, the customer knows that the actual lot will be assigned after he/she arrives at time $(k-1)l$.

In addition, he/she agrees with the system that the assigned parking lot can be at a different location no farther than ε_r^{jk} from the originally reserved lot j. Note that $\varepsilon_r^{jk} \geq 0$ is predetermined by the customer at reservation, which is called the region flexibility. Meanwhile, the customer can tolerate a maximum waiting time to be assigned after his/her reserved parking duration starts, denoted by $\varepsilon_t^{jk} \in [0, l)$. Note that ε_t^{jk} is also predetermined by the customer at reservation, called the time flexibility. We simply denote $\boldsymbol{\varepsilon_r} = \left\{\varepsilon_r^{jk}\right\}$ and $\boldsymbol{\varepsilon_t} = \left\{\varepsilon_t^{jk}\right\}$. It is direct that if customers agree with large location and time flexibility in reservation, the system assignment decision becomes easy to make. However, given this predetermined flexibility, there are still chances when no available parking lot is available to assign, i.e., all parking lots within a radius of ε_r^{jk} from lot j are still occupied until time $(k-1)l + \varepsilon_t^{jk}$. We say the parking reservation service fails in this scenario. The main reason for service failure is the random late departures of previous parking customers. To quantify the impact of late departures, we assume the parking duration of customers are subject to an identical independent distribution, with a mean of $\bar{l}$, where $\bar{l} < l$. Noting that when $\varepsilon_r^{jk} = \varepsilon_t^{jk} = 0$, it indicates that the customer will park exactly at the lot j at time $(k-1)l$, which reduces to a traditional parking reservation service. For the demonstrative purpose, in the following context, we will assume customers who reserved parking lot j during the entire planning horizon $\mathcal{T}$ share the same parking behavior, i.e., sharing the same tolerance on location and time flexibility. Then the index k can be omitted. Later in Section 5.3, we will extend our work to the case of heterogeneous demand.

3.2. Customer parking behavior

In fact, accepting region and time flexibility at reservation incurs disutility for customers. As we consider customers who reserved at the same parking lot over $\mathcal{T}$ at lot j share the same parking preferences, they will jointly react to the parking pricing scheme of the system. To this end, let the disutility of customers who reserved lot j with respect to the region and time flexibility be captured by non-decreasing convex functions $D_r^j(\varepsilon_r)$ and $D_t^j(\varepsilon_t)$, respectively. To encourage the flexible reservation service, the system chooses to compensate customers in the form of price discounts, characterized as p_r^j and p_t^j per unit region and time flexibility, respectively. Therefore, customers who reserve lot j determine their region and time flexibility to minimize their disutility, which is expressed as the following Customer Problem (CP),

$$\min_{\varepsilon_r^j, \varepsilon_t^j} \ \omega^j = p^j - p_r^j \varepsilon_r^j - p_t^j \varepsilon_t^j + D_r(\varepsilon_r^j) + D_t(\varepsilon_t^j) + c \tag{1}$$

$$\text{s.t. } \varepsilon_t^j \in [0, l], \varepsilon_r^j \geq 0, \tag{2}$$

where p^j is the base price at lot j and c is the fixed parking cost such as the in-transit travel cost (Arnott, 2014).

The parking disutility affects not only the flexibility decisions of customers, but also the total amount of parking demand. Suppose $\beta^j \in [0, 1]$ is the reservation occupancy at lot j. It indicates the total parking demand over $\mathcal{T}$ to be $n\beta^j$ at lot j. We consider the short-term problem where individual driver's parking needs (where to park and when to park) will not shift based on the parking price, but the total volume of parking demand at specific locations and time will change over the parking disutility. This assumption is practical when the system pricing scheme is not significantly varying over lot locations. Therefore, to capture the impact of disutility on the demand, we consider the reservation occupancy decreases linearly over the disutility,

$$\beta^j = \beta_0^j - \delta^j \omega^j, \tag{3}$$

where β_0^j and δ^j are the occupancy potential and the unit occupancy reduction, respectively.

Therefore, given the pricing scheme of reservation, i.e., p^j, p_r^j, p_t^j, each slot has a chance β^j to be reserved, with the corresponding customer's flexibility decisions $\varepsilon_r^j, \varepsilon_t^j$. We can use a Bernoulli distribution to capture this. In particular, suppose $a^{jk} = 1$ indicates that the slot $[(k-1)l, kl]$ at lot j is reserved, and $a^{jk} = 0$, otherwise. Then we get $\mathbb{P}(a^{jk} = 1) = \beta^j$, and $\mathbb{P}(a^{jk} = 0) = 1 - \beta^j$. Similarly, we define $\mathbf{a} = \left\{a^{jk}\right\}$.

3.3. System assignment policy

In this section, we establish an evaluation framework to quantify the impact of flexibility on the service successful rate. Given the reservation requests $\mathbf{a}$ and their region and time flexibility $(\boldsymbol{\varepsilon_r}, \boldsymbol{\varepsilon_t})$, the system needs to set up a real-time assignment policy to finally secure a parking slot for each arriving customer. Considering there are cases when the system fails to assign due to late departure customers even with flexibility, a straightforward objective of the policy is to maximize the total number of successful assignments. As the departure time of customers are random, we establish a multi-stage stochastic model.

To capture the real-time dynamics, we assume assignment decisions are made at the beginning of consecutive decision cycles in a much higher frequency than parking reservation, i.e., at time $0, \tau, 2\tau, \ldots,$ where $\tau \ll l$ is the length of each cycle. Hence, the planning horizon $\mathcal{T}$ is further discretized into $\frac{l}{\tau}n$ intervals, or stages, indexed by $\mathcal{S} = \{1, \ldots, \frac{l}{\tau}n\}$. Suppose $x_{ijs} = 1$ if customer i is assigned to lot $j \in J$ at stage $s \in \mathcal{S}$, and $x_{ijs} = 0$, otherwise. Let $\boldsymbol{y}_s$, $s \in \mathcal{S}$, be the state variable at stage s, which is a vector describing the state of the system, e.g., slot availability. Let ξ^s represent the realization of departure uncertainty at the end of stage s, and $\xi^{[1,s]} = \left\{\xi^1, \xi^2, \ldots, \xi^s\right\}$ indicate the sample path of uncertainty realization from beginning to stage s. Similarly, we denote $\mathbf{x} = \left\{x_{ijs}\right\}$ and $\mathbf{y} = \{\boldsymbol{y}_s\}$. Based on this, the system solves the following Assignment Problem ($\text{AP}_{(\mathbf{a},\boldsymbol{\varepsilon_r},\boldsymbol{\varepsilon_t})}$) to maximize the expected total number of successful assignments,

$$\max_{\mathbf{x},\mathbf{y}} A_{(\mathbf{a},\boldsymbol{\varepsilon_r},\boldsymbol{\varepsilon_t})} = \mathbb{E}_\xi \sum_{\{i\}, j\in\mathcal{J}, s\in\mathcal{S}} x_{ijs}\left[\xi^{[1,s]}\right] \tag{4}$$

$$\text{s.t. } \left(\boldsymbol{y}_{s-1}\left[\xi^{[1,s-1]}\right], \boldsymbol{y}_s\left[\xi^{[1,s]}\right], \left\{x_{ijs}\left[\xi^{[1,s]}\right]\right\}\right) \in X^s_{(\mathbf{a},\boldsymbol{\varepsilon_r},\boldsymbol{\varepsilon_t})}, \ \forall \xi^{[1,s]}, s \in \mathcal{S}, \tag{5}$$

$$x_{ijs} \in \{0, 1\}, \ \forall i, \ j \in \mathcal{J}, \ s \in \mathcal{S}, \tag{6}$$

where $A_{(\mathbf{a},\boldsymbol{\varepsilon_r},\boldsymbol{\varepsilon_t})}$ and $X^s_{(\mathbf{a},\boldsymbol{\varepsilon_r},\boldsymbol{\varepsilon_t})}$ indicate the total expected assignments and the corresponding feasible set at stage $s \in \mathcal{S}$, respectively, relying on the reservation requests and corresponding flexibility. Constraints (5) indicate that given the previous system's state $\boldsymbol{y}_{s-1}$, the system assigns customers with decisions $\{x_{ijs}\}$, then with the realization of departure uncertainty ξ^s, the system goes to state $\boldsymbol{y}_s$. Without loss of generality, we assume there is no late departure at the beginning of $\mathcal{T}$ given a sufficiently long rolling horizon, and denote such an initial state by $\boldsymbol{y}_0$. Let $A^*_{(\mathbf{a},\boldsymbol{\varepsilon_r},\boldsymbol{\varepsilon_t})}$ be the optimum assignments under the optimal policy, given the reservation $\mathbf{a}$. Hence, the long-term failure rate of the reservation service under the best assignment policy is

$$q = 1 - \frac{\mathbb{E}_{\mathbf{a}(\boldsymbol{\beta})} A^*_{(\mathbf{a},\boldsymbol{\varepsilon_r},\boldsymbol{\varepsilon_t})}}{n\sum_{j\in\mathcal{J}}\beta_j},$$

where $\mathbb{E}_{\mathbf{a}(\boldsymbol{\beta})} A^*_{(\mathbf{a},\boldsymbol{\varepsilon_r},\boldsymbol{\varepsilon_t})}$ indicates the expected total number of assignments over all possible reservation scenarios given the occupancy levels.

In some special cases, the system-optimal policy may not be in favor of some customers. Such fairness issues are discussed in Jahn et al. (2005). In our system, since the maximum waiting time and maximum detour distance are capped by ε_t and ε_r, respectively, and the service failure rate is restricted to be low, the case of unfairness is limited.

3.4. System pricing policy

With the above building blocks, we now discuss the pricing decision for the system. Since parking is normally a public service, we consider a social surplus (Ss) objective defined as the sum of producer surplus and consumer surplus (Mackowski et al., 2015). In our context, the producer surplus (Ps) is the profit of the system,

$$\text{Ps} = \sum_{j\in\mathcal{J}} n\beta^j\left(p^j - \varepsilon^j_r p^j_r - \varepsilon^j_t p^j_t\right),$$

where $n\beta^j$ is the total demand at lot j and $p^j - \varepsilon^j_r p^j_r - \varepsilon^j_t p^j_t$ is the unit profit of parking service. Meanwhile, the consumer surplus (Cs) can be estimated by the area between the inverse demand curve and the disutility level under equilibrium,

$$\text{Cs} = n\sum_{j\in\mathcal{J}} \int_0^{\beta^j} \left(\frac{\beta^j_0 - u}{\delta^j} - \omega^j\right) du,$$

where the first item in the integrand is the inverse demand curve, directly determined by Eq. (3), while ω^j and β^j are the disutility and occupancy at equilibrium, respectively. Similarly, we define $\mathbf{p} = \left\{p^j\right\}$, $\mathbf{p_r} = \left\{p^j_r\right\}$, and $\mathbf{p_t} = \left\{p^j_t\right\}$. Therefore, we can have the following System's Problem (SP) which maximizes the social surplus,

$$\max_{\mathbf{p},\mathbf{p_r},\mathbf{p_t},\boldsymbol{\varepsilon_r},\boldsymbol{\varepsilon_t},\boldsymbol{\beta}} \text{Ss} = \text{Ps} + \text{Cs} \tag{7}$$

$$\text{s.t. } \boldsymbol{\varepsilon_r}, \boldsymbol{\varepsilon_t} \text{ solves CP}, \ \forall j \in \mathcal{J}, \tag{8}$$

$$\omega^j = p^j - \varepsilon^j_r p^j_r - \varepsilon^j_t p^j_t + D_r(\varepsilon^j_r) + D_t(\varepsilon^j_t) + c, \ \forall j \in \mathcal{J}, \tag{9}$$

$$\beta^j = \beta^j_0 - \delta^j\omega^j, \ \forall j \in \mathcal{J}, \tag{10}$$

$$1 - \frac{\mathbb{E}_{\mathbf{a}(\boldsymbol{\beta})} A^*_{(\mathbf{a},\boldsymbol{\varepsilon_r},\boldsymbol{\varepsilon_t})}}{n\sum_{j\in\mathcal{J}}\beta_j} \le \bar{q}, \tag{11}$$

$$p^j_r, p^j_t, p^j \ge 0, \beta^j \in [0, 1], \ \forall j \in \mathcal{J}. \tag{12}$$

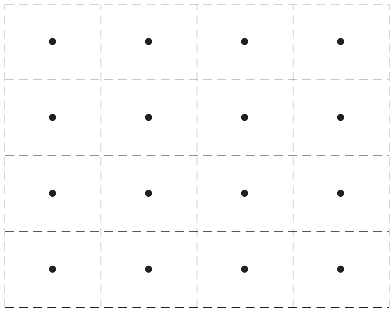


Fig. 1. Lot locations in grid.

Constraints (8) indicate customers optimally choose their flexibility decisions ε_r^j and ε_t^j, given a service pricing scheme p_r^j, p_t^j, and p^j at each lot $j \in \mathcal{J}$. Constraints (9) and (10) enforce the parking demand to satisfy the demand function with respect to the corresponding customer disutility. Constraint (11) quantifies the service failure rate in the long run, where $\bar{q} \in (0, 1)$ is predetermined as the service level threshold. Finally, Constraints (12) set the ranges of decision variables.

It is obvious that SP is a complicated non-convex bilevel optimization problem and NP-hard (Garey and Johnson, 2002), especially given the lower level AP with multi-stage integer decisions. Directly solving SP is impossible for even a moderate-scale instance within reasonable computational time. Therefore, in the following section, we will develop an efficient continuum approximation based solution method and rigorously quantify its optimality gap.

4. Solution algorithm

4.1. Homogeneous problem

We first focus on a homogeneous SP problem. Consider a scenario where all parking lots are evenly distributed on a sufficiently large and simply connected service area $\Omega \subseteq \mathbb{R}^2$ to form a grid with density ρ, i.e., the entire plane is decomposed into identical squares with areas $1/\rho$, and one parking lot is located at the center of each square, as shown in Fig. 1. Sufficiently large Ω guarantees those lots near the boundary have a negligible impact on the solution. In addition, all parking lots share the same parameters, including the customer parking preferences and the demand function. Under such a symmetric setting, the optimal decisions for customers and the system should be spatially indifferent for each lot $j \in J$, except those near the boundary of Ω, which are negligible. We simply remove the superscript j to denote the homogeneous decision p, p_r, p_t, ε_r, ε_t, β. Therefore, we have the following homogeneous SP,

$$\max_{p, p_r, p_t, \varepsilon_r, \varepsilon_t, \beta} n\beta(p - \varepsilon_r p_r - \varepsilon_t p_t) + n\int_0^\beta \left(\frac{\beta_0 - u}{\delta} - \omega\right) du \tag{13}$$

$$\text{s.t. } \varepsilon_r, \varepsilon_t \text{ solves CP,} \tag{14}$$

$$\omega = p - \varepsilon_r p_r - \varepsilon_t p_t + D_r(\varepsilon_r) + D_t(\varepsilon_t) + c, \tag{15}$$

$$\beta = \beta_0 - \delta\omega, \tag{16}$$

$$q(\varepsilon_r, \varepsilon_t, \beta) \le \bar{q}, \tag{17}$$

$$p_r, p_t, p \ge 0, \beta \in [0, 1], \tag{18}$$

where the objective (13) is the average social surplus at each lot, and $q(\varepsilon_r, \varepsilon_t, \beta)$ is the service failure rate given ε_r, ε_t, β. In the following, we will first provide the solution to the homogeneous SP and then extend it to a general case later in Section 4.4.

First, we can rewrite the solution to CP into optimality conditions using Karush-Kuhn-Tucker (KKT) theorem (Kuhn and Tucker, 1951). Given p, p_r, and p_t, we assume $D_r(\cdot)$ and $D_t(\cdot)$ are differentiable. The optimal solution for CP satisfies

$$0 \le \varepsilon_r \perp -p_r + D_r'(\varepsilon_r) \ge 0,$$
$$0 \le \varepsilon_t \perp -p_t + D_t'(\varepsilon_t) + \lambda \ge 0,$$
$$0 \le \lambda \perp l - \varepsilon_t \ge 0,$$

where λ is the corresponding Lagrangian multiplier for constraint $\varepsilon_t \le l$. In most practical cases, $p_r \ge D_r'(0)$ and $p_t \ge D_t'(0)$ hold, which implies that the optimal solution for CP should lie in the interior of the feasible region, i.e., $\varepsilon_r > 0$ and $\varepsilon_t \in (0, l)$. This leads to a simplification of the above complimentary conditions as

$$p_r = D_r'(\varepsilon_r), \quad p_t = D_t'(\varepsilon_t), \tag{19}$$

which replaces Constraints (14).

Next, we address AP, which is relatively difficult. To facilitate the understanding of the problem structure, we discuss the following two-lot problem.

4.2. Two-lot problem

Consider that the homogeneous SP only contains two parking lots that are one unit distance apart. Fortunately, in this two-lot problem, parking assignment becomes very easy so that we can directly obtain the solution to AP. In particular, we consider a policy π as: upon arrival, the customer will be assigned to the reserved slot if it is available, otherwise he/she will be assigned to the earliest available slot within his/her region flexibility. Then we immediately have the following lemma.

Lemma 1. *In a homogeneous two-lot problem, policy π is optimal to AP.*

Proof. When $\varepsilon_r < 1$, no customer accepts the assignment with relocation. Therefore, the problem is reduced to a single lot problem and the system should assign the customer to its reserved lot once it is available. On the contrary, when $\varepsilon_r \ge 1$, all customers accept the assignment for either lot, which is equivalent to a single lot with a capacity of two. So the system should also assign the customer to any available lot(s) as soon as possible. So policy π is the optimal policy in both scenarios. □

We should note that, policy π is optimal mainly due to the symmetric properties. For example, we can easily extend the proof to a homogeneous three-lot problem where the three lots are located at the vertices of an equilateral triangle. However, policy π is not guaranteed to be optimal when customer preferences are heterogeneous.

Now, we can explicitly calculate the long-term failure rate.

Proposition 1. *Given ε_r, ε_t, β, if the parking durations of customers are independent and identically distributed, satisfying an exponential distribution with mean l, the long-run service failure rate of a homogeneous two-lot problem under policy π is:*

$$q = \begin{cases} \left(\frac{1}{2}\beta r_1 + \frac{\varepsilon_t}{l}\beta r_2 + r_2\right)\exp\left(-\frac{2\varepsilon_t}{l}\right), & \varepsilon_r \ge 1, \\ \frac{\bar{l}\beta}{(\bar{l}-\varepsilon_t)\beta + \left(\exp\left(\frac{l}{\bar{l}}\right)-1\right)\bar{l}}\exp\left(-\frac{\varepsilon_t}{l}\right) & \varepsilon_r < 1, \end{cases} \tag{20}$$

where r_1 and r_2 are the solution to a system of equations as shown in Appendix B.

The proof of Proposition 1 is also provided in Appendix B. As a remark, we note that the physical meaning of r_1 and r_2 are the long-run probability of having one and two customers departing late at the beginning of each slot, respectively.

4.3. AP approximation in homogeneous problem

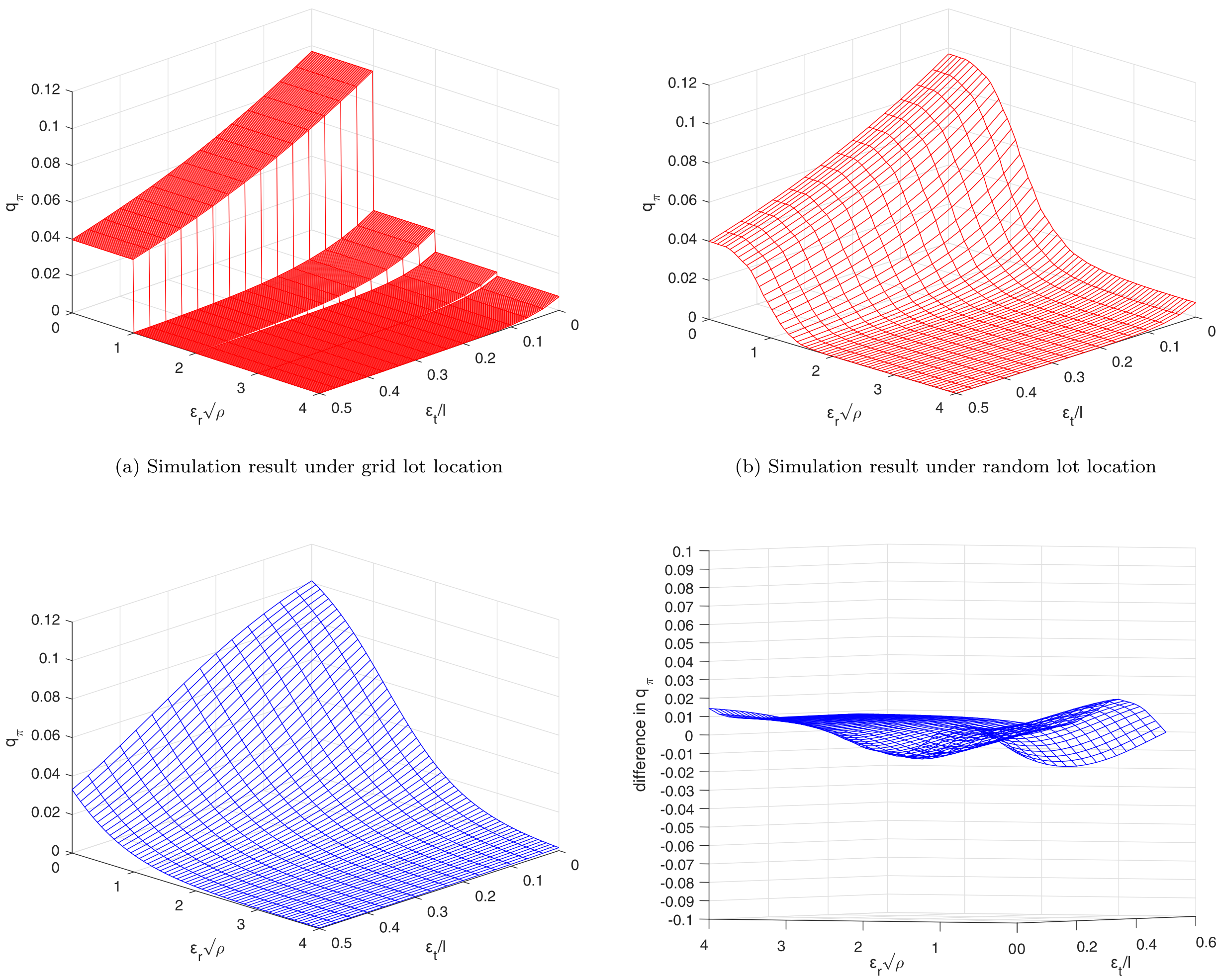
Now we start to extend the result of the two-lot problem into multiple lots in general. Eq. (20) indicates q is a complex nonlinear function with respect to ε_r, ε_t, β. Extension of $q(\varepsilon_r, \varepsilon_t, \beta)$ under multiple homogeneous lots is extremely difficult since π may not be an optimal policy. To this end, we provide an effective way to approximate $q(\varepsilon_r, \varepsilon_t, \beta)$ through Monte-Carlo sampling and preserve analytical tractability. In particular, we simulate to evaluate q under different input values ε_r, ε_t, β, which are further used to "learn" a closed form. However, directly evaluating q requires to solve a huge number of multistage stochastic integer programs, which is not efficient. We note that although policy π is not optimal, it provides a rough estimation to the optimal solution for AP. In addition, evaluating the long-term failure probability under policy π, denoted by $q_\pi(\varepsilon_r, \varepsilon_t, \beta)$, is straightforward in the calculation. This will lead to a feasible solution to AP and $q_\pi(\varepsilon_r, \varepsilon_t, \beta) \ge q(\varepsilon_r, \varepsilon_t, \beta)$ always holds.

Specifically, we suppose $\chi = \left[l^{-1}\varepsilon_t, \sqrt{\rho}\varepsilon_r, \beta\right]$, where l^{-1} and $\sqrt{\rho}$ are introduced to convert the inputs into unit-less variables (Buckingham, 1914). Then we consider the following nonlinear regression,

$$q_\pi = w \cdot \text{sigmoid}(\mathbf{w}_\chi^\intercal \chi + b_\chi) + b + \epsilon_q, \tag{21}$$

where $\text{sigmoid}(\chi) = \frac{1}{1+\exp(-\chi)}$, $w > 0$, and ϵ_q indicates the fitting error. In fact, this is a simple neural network with one hidden layer and a nonlinear sigmoid activation function.

We form a three-dimensional grid of ε_r, ε_t, β and calculate its corresponding q_π to train the network by minimizing the mean squared fitting error. Our training result is shown as Fig. 2. Fig. 2(a) is the simulation result under the grid lot locations. To avoid the jumps on q_π due to discrete relocation distances among symmetric lot locations, we also uniformly sample lot locations under the same density ρ and take the average failure rate over different location samples. Such simulation results and training results are shown in Fig. 2(b) and (c), respectively. And the corresponding training error



(a) Simulation result under grid lot location

(b) Simulation result under random lot location

(c) Training result under random lot location

(d) Training error under random lot location

Fig. 2. Nonlinear regression of q_π.

(input – prediction) is shown in Fig. 2(d). In fact, the domain of most training errors mainly lies in the region where available lots for relocation are close to 0. In a realistic case, a parking lot is seldom isolated. Hence available lots for relocation can be easily above 10 or 20, where our neural network fits the service failure rate quite exactly. To support such arguments, a posterior error estimation is conducted in Section 5.3.

In the following demonstration, we simply approximate $q_\pi \approx w \cdot \text{sigmoid}(\mathbf{w}_\chi^\intercal \chi + b_\chi) + b$, and obtain

$$\mathbf{w}_\chi^\intercal \chi + b_\chi \approx \text{sigmoid}^{-1}\left(\frac{q_\pi - b}{w}\right) := \gamma.$$

Noticing that $\text{sigmoid}^{-1}\left(\frac{q_\pi - b}{w}\right)$ is monotone increasing over q_π, we can equivalently use γ to quantitatively capture the service level. Therefore, given $q_\pi(\varepsilon_r, \varepsilon_t, \beta) \geq q(\varepsilon_r, \varepsilon_t, \beta)$ always holds and ignoring the fitting error, we replace Constraint (17) by

$$\mathbf{w}_\chi^\intercal \chi + b_\chi \leq \bar{\gamma}, \tag{22}$$

where $\bar{\gamma} = \text{sigmoid}^{-1}\left(\frac{\bar{q} - b}{w}\right)$, to yield a tighter feasible region to the homogeneous SP. We denote this problem by $\pi-$restricted homogeneous SP, whose optimal solution is obtained by the following proposition.

Proposition 2. *The optimal interior point solution for the $\pi-$restricted homogeneous SP satisfies the following conditions,*

$$D_t'(\varepsilon_t) = -w_1 l^{-1} \frac{\lambda_\gamma}{\beta}, \tag{23}$$

$$D_r'(\varepsilon_r) = -w_2\sqrt{\rho}\frac{\lambda_\gamma}{\beta}, \tag{24}$$

$$p = \left(w_3\beta - w_1 l^{-1}\varepsilon_t - w_2\sqrt{\rho}\varepsilon_r\right)\frac{\lambda_\gamma}{\beta}, \tag{25}$$

$$D_r(\varepsilon_r) + D_t(\varepsilon_t) = \frac{\beta_0 - \beta}{\delta} - w_3\lambda_\gamma - c, \tag{26}$$

$$\bar{\gamma} - b_\chi = w_1 l^{-1}\varepsilon_t + w_2\sqrt{\rho}\varepsilon_r + w_3\beta, \tag{27}$$

where λ_γ is the scaled Lagrangian multipliers corresponding to Constraint (22), and w_1, w_2, w_3 are scalar elements of $\mathbf{w}_\chi^\intercal$ obtained from the nonlinear regression.

Proposition 2 can be derived from simple algebraic operations on the KKT conditions for the optimal interior point solution of the $\pi-$restricted homogeneous SP, which are provided in Appendix C.

Here we only pay attention to the interior point solution, i.e., $p,\ p_r,\ p_t,\ \varepsilon_r > 0,\ \beta \in (0,\ 1)$ and $\varepsilon_t \in (0,\ l)$ as it captures most realistic scenarios and provide meaningful insights. Proposition 2 provides the necessary conditions for the optimal solution under general customer disutility. Interestingly, we observe that if the customer disutility is linear, the interior point solution may not exist due to Equations (23) and (24). This implies either ε_t or ε_r becomes zero, i.e., customers would not prefer the corresponding flexibility. This is mainly due to the oversimplification of customer disutility, which results in the degeneration of the system into an unrealistic state. To this end, we need to capture the nonlinear characteristic of customer disutility. A natural way is to consider a convex increasing pattern, say, quadratic disutility (Bookbinder and Desilets, 1992), which leads to the following optimality conditions.

Corollary 1. *(i) Assuming $D_t = a_t\varepsilon_t^2$ and $D_r = a_r\varepsilon_r^2$, $a_t, a_r \in \mathbb{R}^+$, the necessary conditions for an interior point solution of the $\pi-$restricted homogeneous SP are*

$$\varepsilon_t = \frac{w_1 l^{-1} a_r}{w_1^2 l^{-2} a_r + w_2^2\rho a_t}\left(\bar{\gamma} - b_\chi - w_3\beta\right),$$
$$\varepsilon_r = \frac{w_2\sqrt{\rho} a_t}{w_1^2 l^{-2} a_r + w_2^2\rho a_t}\left(\bar{\gamma} - b_\chi - w_3\beta\right),$$
$$p = \frac{a_t a_r}{w_1^2 l^{-2} a_r + w_2^2\rho a_t}\left(\bar{\gamma} - b_\chi - w_3\beta\right)^2 + \frac{\beta_0 - \beta}{\delta} - c,$$
$$\lambda_\gamma = \frac{-2n a_t a_r}{w_1^2 l^{-2} a_r + w_2^2\rho a_t}\beta\left(\bar{\gamma} - b_\chi - w_3\beta\right),$$

where β solves

$$\frac{3w_3^2 a_t a_r}{w_1^2 l^{-2} a_r + w_2^2\rho a_t}\beta^2 + \left(\frac{-4w_3 a_t a_r\left(\bar{\gamma} - b_\chi\right)}{w_1^2 l^{-2} a_r + w_2^2\rho a_t} + \frac{1}{\delta}\right)\beta - \frac{\beta_0}{\delta} + c + \frac{a_t a_r\left(\bar{\gamma} - b_\chi\right)^2}{w_1^2 l^{-2} a_r + w_2^2\rho a_t} = 0.$$

(ii) When $-\frac{\beta_0}{\delta} + c + \frac{a_t a_r(\bar{\gamma} - b_\chi)^2}{w_1^2 l^{-2} a_r + w_2^2\rho a_t} < 0$ and $\frac{3w_3^2 a_t a_r}{w_1^2 l^{-2} a_r + w_2^2\rho a_t} + \frac{-4w_3 a_t a_r(\bar{\gamma} - b_\chi)}{w_1^2 l^{-2} a_r + w_2^2\rho a_t} + \frac{1}{\delta} - \frac{\beta_0}{\delta} + c + \frac{a_t a_r(\bar{\gamma} - b_\chi)^2}{w_1^2 l^{-2} a_r + w_2^2\rho a_t} > 0$, there exists a unique solution such that $\beta^* \in (0,\ 1)$.

(iii) The optimal interior point solution exists and is unique when $\beta^* \in (0,\ 1)$, $\varepsilon_t^* \in (0, l)$ and $\varepsilon_r^*, p^*, \lambda_\gamma^* > 0$.

The proof for Corollary 1 can be found in Appendix D.

Corollary 1 provides the sufficient and necessary conditions for the optimal solution of the $\pi-$restricted homogeneous SP under quadratic customer disutility. In realistic cases, δ is normally a very small positive number and β_0 is normally slightly greater than 1, which indicates the conditions in (ii) naturally hold.

Note that the optimal solution of the $\pi-$restricted homogeneous SP generates a feasible solution to the original homogeneous SP and therefore provides a lower bound on the optimal objective of the homogeneous SP.

4.4. Heterogeneous problem

In this section, we extend the algorithm in Section 4.3 to the general case where parameters for parking lots are heterogeneous. To be specific, we consider a large continuous city area $\Omega \in \mathbb{R}^2$, where the lot density $\rho(x)$, reservation length $l(x)$, parameters for the demand function $\beta_0(x)$ and $\delta(x)$, and customer disutility $D_r(\varepsilon_r,\ x)$, $D_t(\varepsilon_t,\ x)$ are slow varying functions with respect to location $x \in \Omega$. Therefore, the optimal solution near location x can be pointwisely approximated by the solution to a corresponding homogeneous problem with local parameters. By extending Proposition 2 to each location $x \in \Omega$, we can obtain the optimal solution for the following $\pi-$restricted heterogeneous SP,

$$\begin{aligned}
\max_{p(x), p_r(x), p_t(x), \varepsilon_r(x), \varepsilon_t(x), \beta(x)} \mathrm{Ss}_\pi = &\int_{x\in\Omega} \rho(x)\Big\{ n\beta(x)(p(x) - \varepsilon_r(x)p_r(x) - \varepsilon_t(x)p_t(x)) \\
&+ n\int_0^{\beta(x)} \left(\frac{\beta_0(x) - u}{\delta(x)} - \omega(x)\right)du\Big\}dx \\
\text{s.t. } &\varepsilon_r(x), \varepsilon_t(x) \text{ solves CP near } x, \\
&\omega(x) = p(x) - \varepsilon_r(x)p_r(x) - \varepsilon_t(x)p_t(x) + D_r(\varepsilon_r, x) + D_t(\varepsilon_t, x) + c(x), \\
&\beta(x) = \beta_0(x) - \delta(x)\omega(x), \\
&w_1 l(x)^{-1}\varepsilon_t(x) + w_2\sqrt{\rho(x)}\varepsilon_r(x) + w_3\beta(x) + b_\chi \le \bar{\gamma}, \\
&p_r(x), p_t(x), p(x) \ge 0, \beta(x) \in (0, 1).
\end{aligned}$$

Similarly, the solution to the $\pi-$restricted heterogeneous SP is a feasible solution to the original SP due to the slow varying assumption. For example, the pricing decision at lot j is simply determined as $p^j = p(x^j)$, where x^j is the location of j. Therefore, Ss_π^* is a lower bound on Ss^*. In the following section, we solve a relaxed problem of SP to obtain an upper bound on the objective and calculate the optimality gap.

4.5. Relaxed SP and optimality gap

In this section, we obtain an upper bound on the optimal objective of SP through relaxation. Under simple algebraic operations, Constraint (11) can be written as

$$(1-\bar{q})n\sum_{j\in\mathcal{J}}\beta_j \le \mathbb{E}_{\mathbf{a}(\boldsymbol{\beta})}A^*_{(\mathbf{a},\boldsymbol{\varepsilon_r},\boldsymbol{\varepsilon_t})},$$

where on the left-hand side is the minimum number of customers that have to be assigned under the service threshold $\bar{q}$. We relax it by penalizing the violation of the constraint in the objective and define the penalty (Pe),

$$\mathrm{Pe} = \mu\left((1-\bar{q})n\sum_{j\in\mathcal{J}}\beta_j - \mathbb{E}_{\mathbf{a}(\boldsymbol{\beta})}A^*_{(\mathbf{a},\boldsymbol{\varepsilon_r},\boldsymbol{\varepsilon_t})}\right),$$

where the constant $\mu > 0$ is the penalty corresponding to each unserved customer. We will provide an algorithm to obtain μ in the later part of this section.

The relaxed SP can then be defined as follows:

$$\max_{\mathbf{p},\mathbf{p_r},\mathbf{p_t},\boldsymbol{\varepsilon_r},\boldsymbol{\varepsilon_t},\boldsymbol{\beta}} \mathrm{Ss_u} = \mathrm{Ps} + \mathrm{Cs} - \mathrm{Pe} \tag{28}$$

$$\text{s.t. } p_r^j = \frac{dD_r^j(\varepsilon_r)}{d\varepsilon_r},\ p_t^j = \frac{dD_t^j(\varepsilon_t)}{d\varepsilon_t},\ \forall j \in \mathcal{J}, \tag{29}$$

$$(9), (10), (12). \tag{30}$$

Relaxed SP is again a problem with multistage integer decisions and therefore can not be solved directly. We further relax this problem by assuming perfect information, i.e., departure uncertainties of all customers are realized at the beginning of the planning horizon. In the scenario based formulation of the relaxed SP, for each $\mathbf{a}(\boldsymbol{\beta})$, multiple departure scenarios are sampled, and one set of decision $\mathbf{p}$, $\mathbf{p_r}$, $\mathbf{p_t}$, $\boldsymbol{\varepsilon_r}$, $\boldsymbol{\varepsilon_t}$, $\boldsymbol{\beta}$ has to be made to maximize the expected objective among all scenarios. Under the assumption of perfect information, instead, the system can make decisions for each scenario, which means the problem becomes separable among scenarios. Therefore, solving the problem is equivalent to solving multiple single-scenario problems. In the following we explicitly formulate the deterministic single-scenario problem.

Denote $\mathcal{I}$ as the set of customers under full reservation, which indicates $|\mathcal{I}| = n|\mathcal{J}|$. Denote $\mathcal{I}_\mathbf{a}$ as the set of customers having a reservation in the system given $\mathbf{a}$, where $\mathbf{a}$ is a sample of reservations given $\boldsymbol{\beta}$ as stated in Section 3.2. For each single-scenario problem, Pe is reduced to

$$\mathrm{Pe} = \mu\left((1-\bar{q})n\sum_{j\in\mathcal{J}}\beta_j - \sum_{i\in\mathcal{I}_\mathbf{a}, j\in\mathcal{J}, s\in\mathcal{S}} x_{ijs}\right).$$

Denote t_i^r, t_i^d, and d_i, $i \in \mathcal{I}$ as the reservation beginning stage, the departure stage, and the reserved parking lot of customer i, respectively. To reduce the number of variables, we define $S_i := \{t_i^r, t_i^r+1, \dots, t_i^d\}$ as the set of stages when customer i is "staying in" the system, where "staying in" indicates the customer has arrived in the parking system and hasn't left.

Adjusting the notation in AP, let $x_{ijs} = 1$ if customer i is assigned to lot $j \in J$ at stage $s \in \mathcal{S}_i$, and $x_{ijs} = 0$, otherwise. Similarly, let $y_{ijs} = 1$ if customer i is parking at lot $j \in J$ at stage $s \in \mathcal{S}_i$, $y_{ijs} = 0$ otherwise. Further denote $\mathcal{I}_s = \{i \in \mathcal{I} \mid s \in \mathcal{S}_i\}$ as the set of customers "staying in" the system at stage s, and $d \in \mathbb{R}_+^{|\mathcal{J}|\times|\mathcal{J}|}$ as the distance matrix between any two locations. Therefore the deterministic Single-Scenario Problem (SSP) can be defined as follows:

$$\max_{\mathbf{p},\mathbf{p_r},\mathbf{p_t},\boldsymbol{\varepsilon_r},\boldsymbol{\varepsilon_t},\boldsymbol{\beta},\mathbf{x},\mathbf{y}} \mathrm{Ss_u} = \mathrm{Ps} + \mathrm{Cs} - \mathrm{Pe} \tag{31}$$

s.t. (29), (9), (10), (12),

$$x_{ijs} d_{d_i j} \le \varepsilon_r^{d_i}, \ \forall i \in \mathcal{I}, j \in \mathcal{J}, s \in \mathcal{S}_i, \tag{32}$$

$$s - \left(t_i^r + \varepsilon_t^{d_i}\right) + \frac{l}{\tau}\left(x_{ijs} - 1\right) \le 0, \ \forall i \in \mathcal{I}, j \in \mathcal{J}, s \in \mathcal{S}_i, \tag{33}$$

$$\sum_{j \in \mathcal{J}, s \in \mathcal{S}_i} x_{ijs} \le 1, \ \forall i \in \mathcal{I}, \tag{34}$$

$$x_{ijs} \le y_{ijk}, \ \forall i \in \mathcal{I}, j \in \mathcal{J}, s \in \mathcal{S}_i, k \in \left\{s, s+1, \ldots, t_i^d\right\}, \tag{35}$$

$$x_{ijs} + y_{ij(s-1)} \ge y_{ijs}, \ \forall i \in \mathcal{I}, j \in \mathcal{J}, s \in \mathcal{S}_i \cap \left\{2, 3, \ldots, \frac{l}{\tau} n\right\}, \tag{36}$$

$$x_{ij1} \ge y_{ij1}, \ \forall i \in \mathcal{I} : 1 \in \mathcal{S}_i, j \in \mathcal{J}, \tag{37}$$

$$\sum_{i \in \mathcal{I}_s} y_{ijs} \le 1, \ \forall j \in \mathcal{J}, s \in \mathcal{S}, \tag{38}$$

$$x_{ijs} \in \{0,1\}, y_{ijk} \in \{0,1\}, \ \forall i \in \mathcal{I}, j \in \mathcal{J}, s \in \mathcal{S}_i, k \in \mathcal{S}_i \cup \{\max(\min(\mathcal{S}_i) - 1, 1)\}. \tag{39}$$

Constraints (32) indicate that any feasible assignment has to be within the customer's region flexibility. Constraints (33) indicate that any feasible assignment must happen before a customer waiting for more than the length of the time flexibility, i.e., if $x_{ijs} = 1$, then $s \le \left(t_i^r + \varepsilon_t^{d_i}\right)$. Constraints (34) enforce that each customer can at most be assigned once. Constraints (35) indicate that if customer i is assigned to lot $j \in J$ at stage $s \in \mathcal{S}_i$, then he/she occupies lot j until departure. Constraints (36) indicate that if customer i is parking at lot $j \in J$ at stage $s \in \mathcal{S}_i$, then either he/she was already at lot j at stage $s-1$, or he/she starts parking at stage s. Constraints (37) are a special case of Constraints (36), for the first stage of the planning horizon. Constraints (38) are the capacity constraints. Constraints (39) set the ranges of some decision variables.

The value of μ is important because it influences the gap between the optimal objective of SP and the relaxed SP. In the following we propose Algorithm 1 to estimate μ. The idea of the algorithm is to find the μ value such that Pe is close to 0, indicating no violation of the relaxed service constraint (Fisher, 1981). We take the average consumer surplus under the optimal solution of the $\pi-$restricted homogeneous SP as the starting point, and iteratively update μ based on the gradient information until meeting the termination condition, which is set as the ratio of the absolute value of the penalty to the objective has to be less than a given threshold ϵ for consecutive three iterations. In the algorithm, we denote β^* as the optimal solution of the $\pi-$restricted homogeneous SP, and ω^* as the corresponding customer disutility.

Algorithm 1 Determining μ.

Initialization: set $\mu = \frac{\int_0^{\beta^*}\left(\frac{\beta_0 - u}{\delta} - \omega^*\right)du}{\beta^*}$, $K = 0$, $M = 0$, and $\epsilon > 0$.
while $M < 3$ do
  $K \leftarrow K + 1$
  solve SSP(μ), and denote the resulting objectiveand penalty as $\mathrm{Ss_u^*}$ and $\mathrm{Pe^*}$ respectively.
  if $\frac{|\mathrm{Pe^*}|}{\mathrm{Ss_u^*}} < \epsilon$then
    $M \leftarrow M + 1$
  else then
    $\mu \leftarrow \mu + K^{-1/3}\frac{\mathrm{Pe^*}}{\mu}$
    $M \leftarrow 0$
  end if
end while
return μ

With the relaxation procedures above, the expectation of the optimal objective values of SSP is greater than the optimal objective of SP, i.e., $\mathbb{E}[\text{Ss}_\text{u}^*] \geq \text{Ss}^*$. Due to the nonlinearity of $D_r(\varepsilon_r)$ and $D_t(\varepsilon_t)$, SSP is unsuitable for commonly used MIP solvers such as CPLEX and Gurobi. Therefore, further tunings on the model using existing relaxation techniques are necessary. In Appendix E, we provide the full details of such tuning steps, including piecewise McCormick envelope and the relaxation based on Difference of Convex functions (D.C.) decomposition, for the case of quadratic $D_r(\varepsilon_r)$ and $D_t(\varepsilon_t)$. We denote the objective of the final version of the problem as $\hat{\text{Ss}}_\text{u}$. The fact that $\hat{\text{Ss}}_\text{u}^*$ is an overestimation for Ss_u^*, along with the result in Section 4.4 imply that $\text{Ss}_\pi^* \leq \text{Ss}^* \leq \mathbb{E}[\text{Ss}_\text{u}^*] \leq \mathbb{E}\left[\hat{\text{Ss}}_\text{u}^*\right]$. Therefore we conclude that a feasible solution for SP can be obtained by solving the $\pi-$restricted SP, and the resulting objective value is bounded with an optimality gap of $\frac{\mathbb{E}\left[\hat{\text{Ss}}_\text{u}^*\right] - \text{Ss}_\pi^*}{\mathbb{E}\left[\hat{\text{Ss}}_\text{u}^*\right]}$.

5. Numerical study

In this section, we first study the effects of parameters on the optimal interior point solution of the $\pi-$restricted homogeneous SP. Then we compare the flexible parking reservation system with its non-flexible counterpart via a hypothetical case study of six lots. Following that, a case study of a large urban area of San Francisco is conducted.

We first introduce some parameters and settings that are used through the section. We set the length of the horizon, the length of each reservation, and the length of each decision circle to be $n = 9$ h, $l = 60$ min, and $\tau = 3$ min, respectively. Addition to the assumption of i.i.d. customers' parking time, for the sampling purpose, we adopt the exponential distribution, with the probability of the late departure set at 15%. The service failure threshold is set to be $\bar{q} = 1\%$. To study the impact of parking pricing on customers' parking behavior, we need to consider the in-transit travel cost generated in parking, which can be measured by the time value of the commute time users spend on downtown streets (Arnott et al., 2015). We set travel distance to be 2.0 miles, with the average cruising speed of 10 mph, and time value of \$15.6 per hour (Mackowski et al., 2015). Based on that, the in-transit travel cost for each customer is $c = \$3.12$. With the assumption that the demand is linearly influenced by the disutility, with elasticity e, we can estimate the demand function as

$$e\frac{\omega - (p' + c)}{p' + c} = \frac{\beta - \beta'}{\beta'}, \tag{40}$$

where p' and β' denote the average parking rate and the average occupancy in peak hours, respectively. Based on Eq. (40), parameters in the demand function $\beta = \beta_0 - \delta\omega$ are given as $\beta_0 = \beta'(1 - e)$ and $\delta = -\frac{e\beta'}{p'+c}$. For the sampling purpose, we assume the reservation requests at each lot in each time interval are independent. The disutility functions are again assumed to be $D_t = a_t\varepsilon_t^2$ and $D_r = a_r\varepsilon_r^2$. We estimate $a_t = \$15.6\ \text{h}^{-2}$ based on the value of time, and estimate $a_r = \$1\ (100\,\text{m})^{-2}$ according to the following consideration. The time value of walking for 100 m is \$0.3234, with the average walking speed being 1.34 m/s (Hoogendoorn and Bovy, 2004). Considering other disutility related to the pedestrian walking such as discomfort (Hoogendoorn and Bovy, 2004), we set $a_r = \$1\ (100\,\text{m})^{-2}$. The threshold for the termination of μ estimation is set as $\epsilon = 0.003$.

5.1. Effects of parameters

In this section, we study the effects of parameters on the outcome of the flexible reservation system. Numerical tests are conducted to show how the solution p, p_r, p_t, ε_r, ε_t, β, given by Corollary 1, changes with the perturbation on input parameters. We further study the effects of parameters on economical measurements Ss_π, ω, and $P = p - 2a_r\varepsilon_r^2 - 2a_t\varepsilon_t^2$, i.e., social surplus of each parking lot, individual user's disutility, and the out-of-pocket fee of each parking reservation.

In the benchmark case, we set $e = -0.3$ (TCRP, 2005), $p' = \$2.58$/h, $\beta' = 80\%$ (SFMTA, 2014a) and $\rho = 6721$ mile^{-2}, which is the average lot density obtained from SFMTA (2018). Then we obtain $\beta_0 = 1.04$ and $\delta = 0.042$ \$$^{-1}$. Under such parameter settings, the optimal interior point solution of the $\pi-$restricted homogeneous SP and its corresponding economical measurements can be obtained as shown in Table 2. We then perturb each parameter from the benchmark value while keeping others unchanged. The implications of parameters to decision variables and economical measurements in the equilibrium are also summarized in Table 2.

Table 2
Effects of parameters.

Para.	β	p \$/h	p_r \$/(100 m)^2	ε_r m	p_t \$/(10 min)^2	ε_t min	ω \$	P \$	Cs_π \$	Ps_π \$	Ss_π \$
	84.50%	2.02	0.086	69.08	0.041	2.84	4.63	1.00	76.31	7.61	83.92
ρ ↑	↑	↓	↓	↓	↓	↓	↓	↓	↑	↓	↑
a_r ↑	↓	↑	↓	↓	↑	↑	↑	↑	↓	↑	↓
a_t ↑	↓	↑	↑	↑	↓	↓	↑	↑	↓	↑	↓
δ ↑	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
β_0 ↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑

Note: In the table, ↑ and ↓ indicate the term monotonically increases and decreases, respectively.

Table 3
The optimal solutions under different scenario profiles.

Scen.	β	p \$/h	p_r \$/(100 m)^2	ε_r m	p_t \$/(10 min)^2	ε_t min	ω \$	P \$	Cs_π \$	Ps_π \$	Ss_π \$
1	99.41%	2.37	0.093	75.05	0.041	2.83	4.89	1.17	105.63	10.49	116.12
2	100%	22.59	0.101	81.46	0.048	3.35	25	21.17	225	190.51	415.51

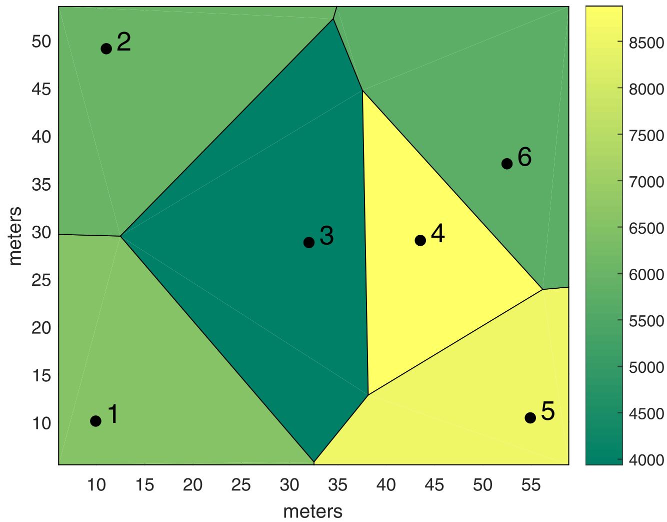


Fig. 3. Voronoi diagram and lot density $\left(\text{mile}^{-2}\right)$.

First, when the lot density ρ increases, customers can share more lots within a smaller region flexibility, thus, a smaller time flexibility is required to guarantee the service level. This allows the system to handle more customers and leads to the drop of the base price. Meanwhile, customers will be better off with a lower disutility. As for the surpluses, both customer and social surpluses increase despite a slight decrease in producer surplus. Second, with the increase of a_r (or a_t), customers are less tolerant to the region (or time) flexibility, which needs to be compensated by a higher time (or region) flexibility to meet the service requirement. Consequently, the service prices rise, together with a higher customer disutility. As for the surpluses, on the contrary, both consumer and social surpluses decrease at a slight increase of the producer surplus. Third, when customers' sensitivity towards the disutility δ increases, both demand and disutility drop in equilibrium. Therefore, less flexibility is necessary to meet the service level. All surpluses decrease. Finally, when there are potentially higher demand β_0 in this area, intuitively, higher region and time flexibility are necessary to guarantee the service level. All surpluses increase.

Then we discuss the solutions under some special parameter profiles, as shown in Table 3. Parameters not specified in the scenarios take the same values as in the benchmark case. The first scenario reflects downtown areas during peak hours, where parking demand ($\beta_0 = 1.2$) and lot density ($\rho = 8000$ mile^{-2}) are high. In the solution, we can observe that parking lots are almost fully utilized to satisfy the high demand, and a higher region flexibility combined with the higher lot density enable the system to guarantee the service level. The second scenario represents the parking case under a special event such as a concert or a sport game, where the demand potential ($\beta_0 = 1.2$) is high and inelastic ($\delta = 0.02$ \$$^{-1}$). In the solution, lots are 100% reserved, and the base price is extremely high.

5.2. Flexible parking reservation system vs. non-flexible parking reservation system

In this section we compare the flexible parking reservation system with its non-flexible counterpart, based on a hypothetical case study of six lots with heterogeneous lot density. We first implement the proposed algorithm to the 6-lot problem. Given the locations of six lots, we approximate the density for each based on the inverse of the area it covers, which is determined by a so-called Voronoi diagram (Aurenhammer, 1991) as shown in Fig. 3. In a Voronoi diagram, the region is partitioned into multiple subregions. In each subregion, the distance to its Voronoi center (parking lot) is the minimum (comparing to other lots). We set the values of e, p' and β' to be the same as those in Section 5.1. By implementing the proposed method, we obtain the following solution in Table 4. Results show that the proposed flexible parking reservation system can utilize over 84% parking resources on average with less than 1% service failure rate. The average out-of-pocket fee (\$1.06/h) and average customer disutility (\$4.73) are lower than the values in the SFpark project (\$2.58/h and \$5.70, respectively). The average region and time flexibility are 70.50 m and 3.04 min, respectively, which lie in reasonable ranges. For locations with high parking density, e.g., lot 4 and 5, the optimal parking occupancy is high, and the base price and flex-

Table 4
Solution of 6-lot problem.

Lot index	1	2	3	4	5	6
Parking density ρ (mile^{-2})	6536	5990	3936	8878	8560	5721
Occupancy β^*(%)	84.35	83.88	81.26	85.81	85.65	83.62
Base price p^* ($/ hour)	2.07	2.22	3.05	1.60	1.65	2.30
Discount for region flexibility p_r^* ($/(100 meter)^2)	0.087	0.090	0.103	0.077	0.078	0.091
Region flexibility ε_r^* (meter)	69.80	72.05	82.89	62.04	62.95	73.24
Discount for time flexibility p_t^* ($/(10 minute)^2)	0.042	0.045	0.064	0.032	0.033	0.047
Time flexibility ε_t^* (minute)	2.91	3.13	4.45	2.22	2.29	3.26
Out-of-pocket fee ($)	1.02	1.10	1.51	0.79	0.82	1.14
Customer disutility ω^* ($)	4.67	4.78	5.40	4.32	4.36	4.84
Social surplus Ss_π^* ($)			501.80			
Upper bound on social surplus $\mathbb{E}\left[\hat{Ss_u}^*\right]$ ($)			525.37			
Optimality gap			4.48%			

Table 5
Flexible parking reservation system vs. non-flexible parking reservation system (late departure probability = 15%).

Performance	Benchmark flexible system	Non-flexible parking reservation system					
		Setting I			Setting II		
		$\bar{q}=1\%$	$\bar{q}=2\%$	$\bar{q}=4\%$	$\beta_{\text{non}}=84\%$	$\beta_{\text{non}}=74\%$	$\beta_{\text{non}}=64\%$
Occupancy	84% (average)	56.27%	65.69%	75.17%	84%	74%	64%
Service failure rate	1%	1%	2%	4%	6.35%	3.70%	1.91%

ibility are low, which further indicates low customer disutility. An upper bound on the social surplus is obtained following the procedure covered in Section 4.5, and the optimality gap of 4.48% indicates the solution obtained through the proposed framework is nearly optimal.

We then compute the outcome of the non-flexible parking reservation system. To avoid the service failure, we set that when the reserved lot is not available upon a user's arrival, the system would relocate the user to the closest available lot. Service failure occurs when a customer cannot be assigned to any lot upon arrival. The comparison is conducted in the following two settings.

In Setting I, we restrict the service threshold for the non-flexible system, and the comparison is focused on the utilization of parking resources. We set the service threshold in a sequence of $\bar{q}=1\%, 2\%, 4\%$ and conduct comparison separately. The optimal occupancy for the flexible parking reservation system is provided as $\{\beta^*\}$ in Table 4. For the non-flexible system, a natural objective is to find the maximal occupancy under which the service threshold can be satisfied. For simplicity, we set that all lots in the non-flexible parking system will share the same occupancy level, denoted as β_{non}. Through a bi-section search based on results of discrete simulation, we obtain the maximum feasible occupancy level as β^*_{non}. This uniform setting of occupancy level is a good approximation to a heterogeneous profile from the observation that β^* tend to be very close across different lots. In Setting II, we set the lot reservation occupancy in the non-flexible system in a sequence of 84%, 74%, 64%, and conduct comparison focusing on the service failure rate.

Computational results under two settings are provided in Table 5. Under Setting I, when $\bar{q}=1\%$, i.e., the same as in the flexible benchmark case, the optimal reservation occupancy is only 56.27%, which is far less than $\{\beta^*\}$ whose values are all greater than 81%. Even under a high service failure rate of $\bar{q}=4\%$, the optimal occupancy (75.17%) is still far less than 84%. Under Setting II, when the reservation occupancy equals 84%, the average service failure rate is $q=6.53\%$, which is over six times of the service threshold. Even when the reservation occupancy drops to 64%, the service level still doesn't meet the threshold. The above results indicate that the proposed flexible parking reservation system dominates its non-flexible counterpart.

To further investigate how much the time and region flexibility contribute to reduce the service failure rate individually, we also evaluate the service failure rate when either time or region flexibility is used. Table 6 provides the flexibility profiles, where the flexibility values are taken from Table 4. The resulting service failure rate in Scenario 1 and 2 are 5.78% and 10.38%, respectively, which shows that each flexibility is effective in reducing the chance of service failure, but is still far from satisfying the service threshold. This indicates that by combining moderate time flexibility and region flexibility, the system avoids imposing extremely high value of a single flexibility to customers.

Note that the above results are based on the setting that customers have a 15% chance of late departure. When this value is smaller, i.e., more customers are obeying the rules, we would expect less difference between the flexible system and its non-flexible counterpart. We test such a scenario by setting the late departure percentage at 5% with other parameters unchanged, and the result is shown in Table 7. We observe that under the service threshold $\bar{q}=1\%$, the flexible system enables over 13% more occupancy than the non-flexible system does, which is significant, though much less than the dif-

Table 6
Flexibility profile.

Lot index		1	2	3	4	5	6
Occupancy β(%)		84.35	83.88	81.26	85.81	85.65	83.62
Scenario 1	ε_r (meter)	69.80	72.05	82.89	62.04	62.95	73.24
	ε_t (minute)	0	0	0	0	0	0
Scenario 2	ε_r (meter)	0	0	0	0	0	0
	ε_t (minute)	2.91	3.13	4.45	2.22	2.29	3.26

Table 7
Flexible parking reservation system vs. non-flexible parking reservation system (late departure probability = 5%).

Performance	Benchmark flexible system	Non-flexible parking reservation system					
		Setting I			Setting II		
		$\bar{q} = 1\%$	$\bar{q} = 2\%$	$\bar{q} = 4\%$	$\beta_{\text{non}} = 89\%$	$\beta_{\text{non}} = 79\%$	$\beta_{\text{non}} = 69\%$
Occupancy	89% (average)	75.90%	86.83%	98.24%	89%	79%	69%
Service failure rate	1%	1%	2%	4%	2.24%	1.15%	0.54%

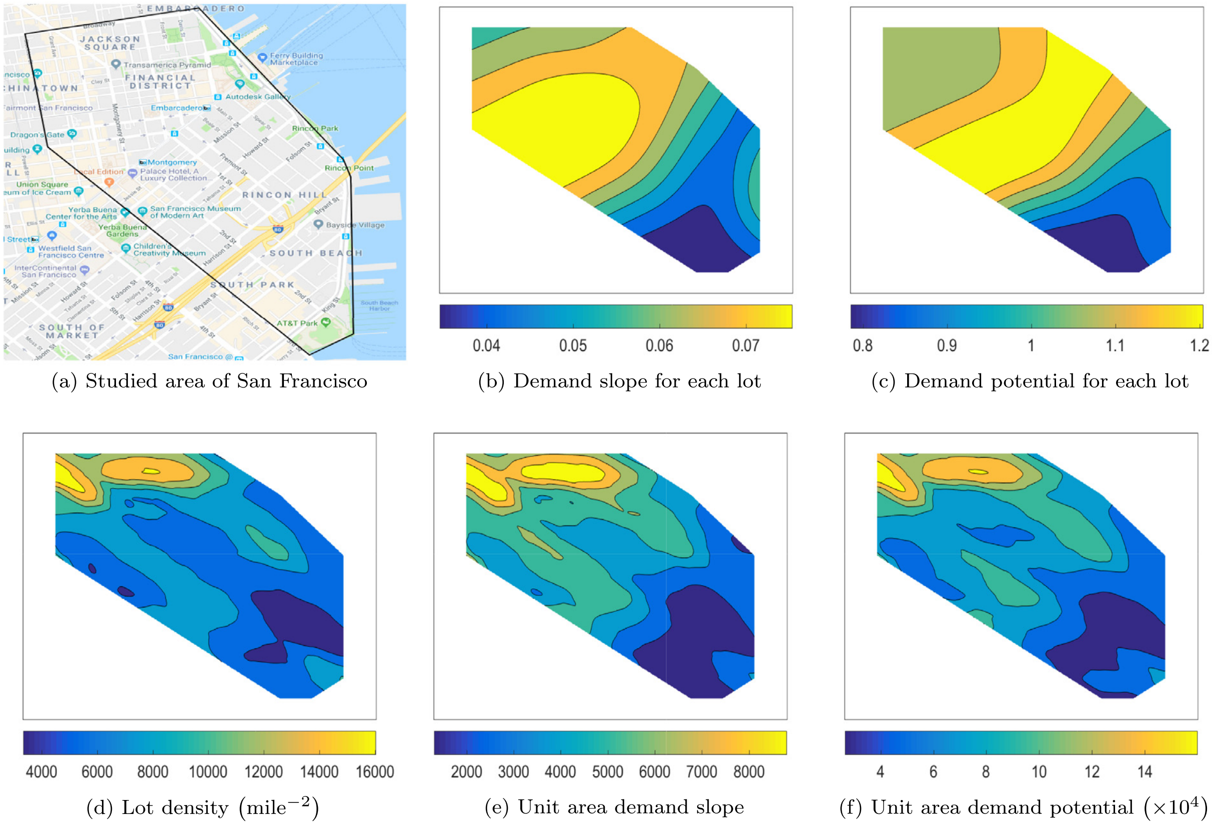
ference (27%) when the chance of late departure is 15%. When $\bar{q}$ increases and the late departure probability is low, it is easier for the non-flexible parking system to reach the same level of occupancy as the flexible parking system. Under Setting II, when the occupancy is set at 79% (10% less than the benchmark value), the service failure probability (1.15%) is already close to the targeting threshold. While when the late departure probability is 15%, a 20% drop in the occupancy (84% to 64%) is not enough to satisfy the service threshold. Based on the facts above, we conclude that the flexible parking reservation system clearly outperforms its non-flexible counterpart even with a relatively small portion of late departing customers, and the benefit of it becomes more significant when customers have a higher chance of late departure.

5.3. Generalization to large-scale problems

In this section, we apply the proposed algorithm to a case study of a selected urban area of San Francisco shown as the polygon area in Fig. 4(a). Data used in this section are from the SFpark project. According to the location data in SFMTA (2018), there are 6,280 parking meters in the selected area. We consider heterogeneous lot density and demand functions. Lot density over the selected area can be estimated based on the location data using the kernel interpolation (Parzen, 1962), and the result is provided in Fig. 4(d). The spatial varying demand functions are estimated based on hourly occupancy and price data of parking meters from 2011 to 2013 (SFMTA, 2014b). To capture the highly congested planning horizon, we use the data recorded during weekday busy hours (12 p.m.– 2 p.m.). We conduct linear regression to estimate the demand function at each location, following which kernel interpolation is used to estimate the demand function over the space. The resulting $\delta(x)$ and $\beta_0(x)$ per lot are shown in Fig. 4(b) and (c), respectively. In addition, the corresponding prorated demand function per unit area is shown in Fig. 4(e) and (f). Following the procedure in Section 4.4, we obtain a feasible solution for the problem as shown in Fig. 5. Before digging into the implications of the solution, we test the posterior error of the nonlinear regression on the service failure rate. Simulations show that the obtained solution results in an average service failure of 1.04%, which is very close to the targeting level of 1%.

With both lot density and demand patterns being heterogeneous, we first study their roles in determining the decisions and economical measurements in the equilibrium. Figs. 4(b), (c) and 5(f) indicate that the occupancy at each lot is mainly determined by its demand potential, where high occupancy is widely seen in areas of large demand potential. We also observe some outliers where high lot occupancy appears in the upper left areas with a relatively low demand potential. This is because customers are less sensitive to the price in this area. Comparing Figs. 4, 5(b) and (c), we find that the lot density, instead of the demand function, plays the key role in determining region and time flexibility. This implies our flexible reservation system is more effective in a congested urban area with highly uneven distributed parking lots. Moreover, large flexibility is also observed in the center areas where lot density is not the smallest but the demand is the highest. This indicates our flexible system contributes more when parking is extremely congested. Last, we have some interesting observations that pricing decisions vary mainly with the pattern of lot density rather than the typical demand function (Figs. 4(b), (c), (d) and 5(a)). In the high lot density areas, low base price, along with low time and region flexibility discounts tend to appear. While in areas of low lot density, all three pricing terms are high. This is due to the introduction of flexibility, which provides additional demand management instruments so that the spatial heterogeneity of parking resources can be better hedged through relocation. The local surge of demand can be absorbed through nearby parking sharing.

Next we study how surpluses (Fig. 6) change under heterogeneous inputs. Comparing Figs. 4(e), (f) and 6(a), we observe that the consumer surplus is mainly determined by the unit-area demand potential. Even in areas of large demand slope, if the demand potential is large, the customer surplus will still have a large value. The unit-area producer surplus (Fig. 6(b))



(a) Studied area of San Francisco (b) Demand slope for each lot (c) Demand potential for each lot

(d) Lot density (mile^{-2}) (e) Unit area demand slope (f) Unit area demand potential $(\times 10^4)$

Fig. 4. Heterogeneous inputs in the San Francisco case study.

shares the similar pattern to the lot occupancy (Fig. 5(f)), and therefore is mainly determined by the demand pattern. As for the social surplus (Fig. 6(c)), results show that the consumer surplus accounts for a major portion of it, and they vary similarly over space. Note that the balance between the consumer surplus and the producer surplus can be easily adjusted by changing their weights in the objective function. Therefore, the proposed model can fit the purpose of different types of parking management agencies, from the public service provider (as shown in the context of this paper) to a private agency (by potentially setting zero weight to the consumer surplus).

Now we discuss the incentives to implement this flexible parking reservation system, from the perspective of customers and the parking management agency. Again, the data used here are collected during weekday busy hours (12 p.m.– 2 p.m.). For customers, if compared to the average price in the SFpark project shown in Fig. 7(a), the out-of-pocket fee (Fig. 5(d)) is smaller, with most values lying under $1.2/hour. Even when considering the disutility caused by flexibility, the customer disutility (Fig. 5(e)) is also much smaller than the average disutility in the SFpark project (Fig. 7(b)). Therefore the flexible parking system will benefit customers by providing secured parking services under a lower disutility. From the standpoint of the parking management agency, the utilization of parking resources is improved to a large extent. Fig. 7(c) shows the average occupancy over the studied area, under the dynamic price adjustments of SFpark. It is clear that the occupancy under the proposed system significantly outperforms the value in SFpark. We also compare the flexible parking reservation system with the traditional parking reservation system when waiting or relocation is not allowed. Results show that under the same occupancy as shown in Fig. 5(f), the traditional system returns an average service failure of 11.29%, which is much higher than the 1% threshold. One may have doubts about the incentives for a parking management agency to implement the proposed system since the producer surplus seems small. However, the public service provider (e.g., Department of Transportation) focuses more on the overall performance of the city's traffic network. The flexible reservation system eliminates the traffic cruising for parking, which means the congestion in urban areas can be relieved to a large extent, which further indicates a significant improvement in the efficiency of the society. Even for the private parking management agencies, coefficients in the system's objective can be adjusted to focus on generating profits.

We then clarify when the flexible parking reservation system should be implemented. Such a system is proposed to tackle the service failure in traditional parking reservation systems caused by uncertainty in customers' departing behavior. In areas where parking is easy, there is no need for parking reservation, and thus it is unnecessary to implement any reservation-based parking system. In busy downtown areas, if most customers are well-behaved and follow the reservation

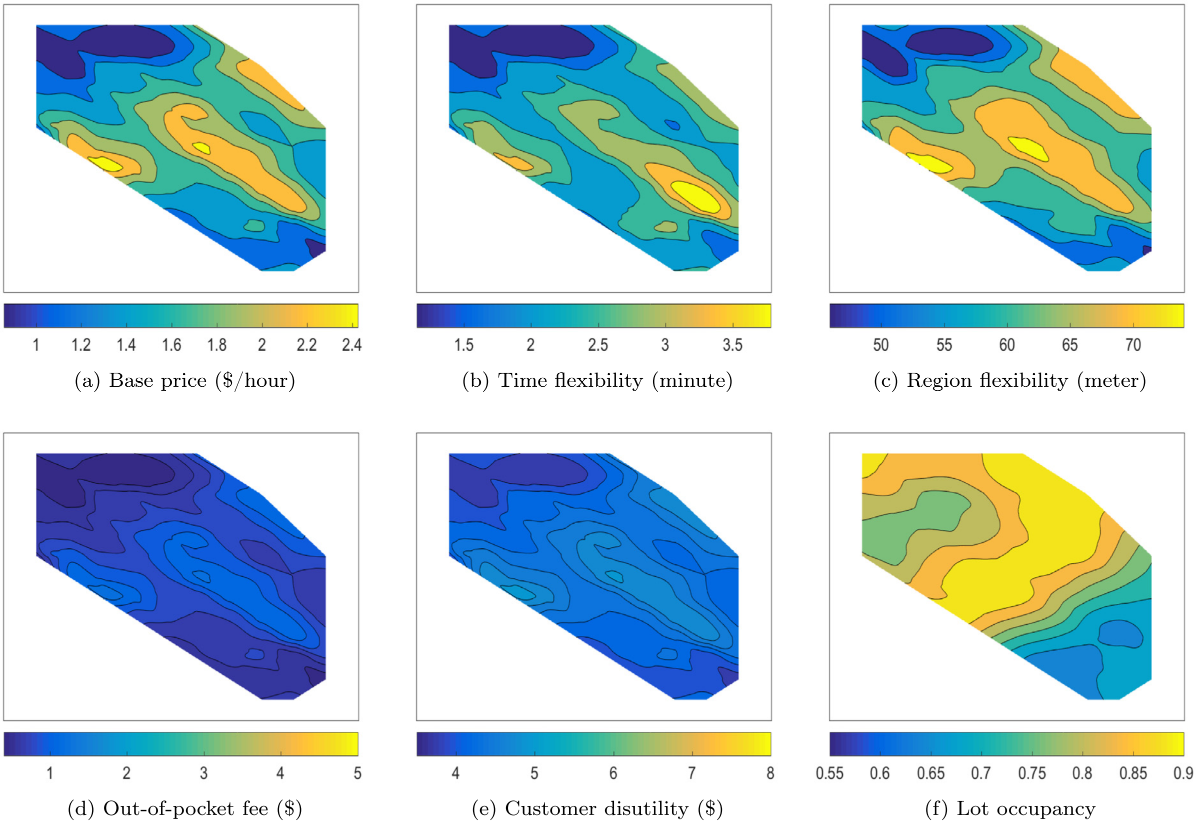


Fig. 5. Outcome: prices, occupancy and flexibility.

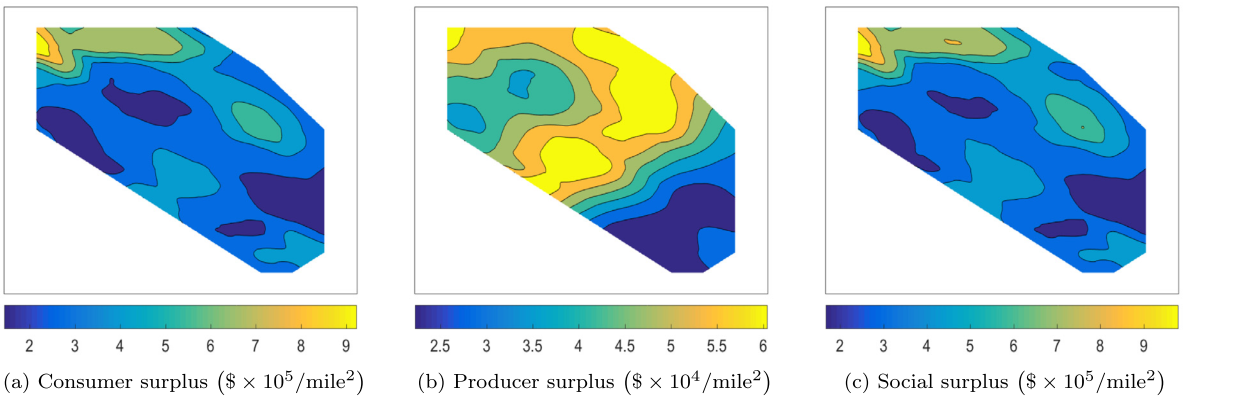
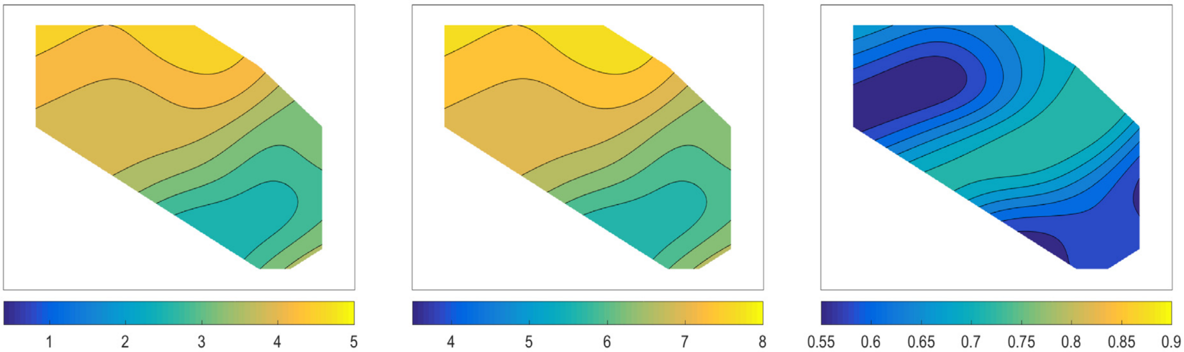


Fig. 6. Outcome: surpluses.

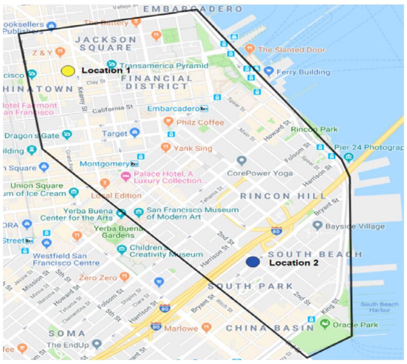
schedule, then it is better to maintain the traditional pricing strategy. Another scenario unsuitable for the proposed system is when the parking length is long. Under such a circumstance, the impact of late departures can usually be neglected. Thus, it is unnecessary to introduce flexibility into the system.

Next we study problems with time-varying demand functions. CA is used again to approximate the optimal solutions. To be specific, the optimal solution near time t can be point-wisely approximated by the solution to a corresponding homogeneous problem with parameters valued at t. In the following we implement above procedures to the San Francisco case study. For ease of representation, we select two locations as shown in Fig. 8(a), with the lot density near location 1 and 2 being 17079 mile^{-2} and 4934 mile^{-2}, respectively. Based on the data file (SFMTA, 2014b), we estimate the demand functions for the two locations at discrete time points during weekdays 7 a.m.– 5 p.m., and then approximate the demand

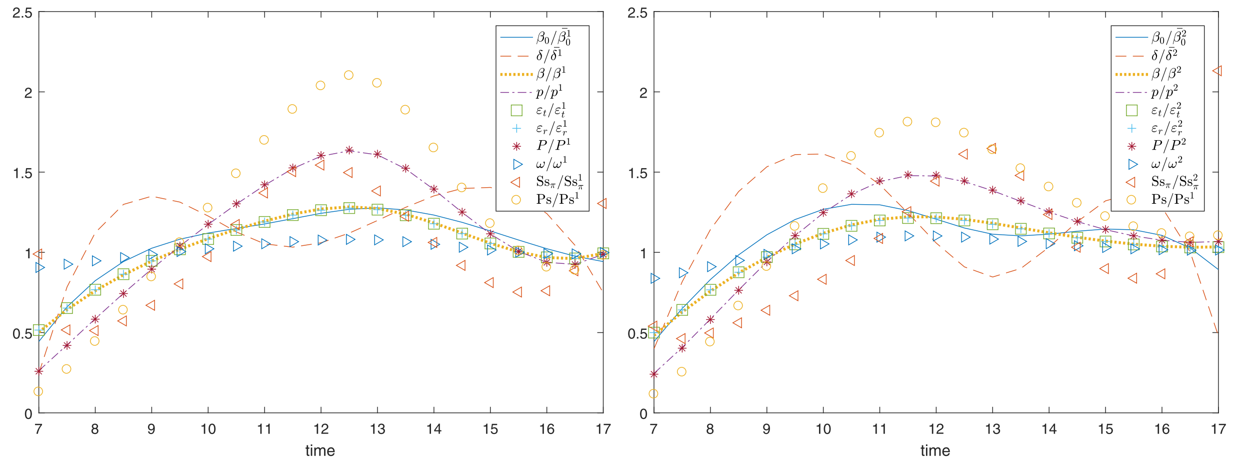


(a) SFpark: Average price ($/hour) (b) SFpark: Average customer disutility ($) (c) SFpark: Average occupancy

Fig. 7. Outcome of SFpark.



(a) Selected locations



(b) Location 1: relative variations of the parameters and the solution (c) Location 2: relative variations of the parameters and the solution

Fig. 8. Selected locations, demand functions, and solutions.

functions via interpolation. We then obtain the feasible solutions via CA. To present the results, we first provide the benchmark solutions in Table 8 when both parameters for the demand function take their mean values, denoted as $\bar{\beta}_0$ and $\bar{\delta}$. Then we draw the ratio graphs to depict how solutions change with the varying demand, as shown in Fig. 8(b) and (c), where superscripts 1 and 2 are used to indicate the benchmark value for location 1 and location 2, respectively. We observe that pricing terms reach their peaks in the middle of the day, when the demand potential is relatively large and the demand slope is small. This is because the demand (occupancy) has to be controlled so that the service threshold can be satisfied.

Table 8
Benchmarks under the mean demand function.

Lot	Input		Solution											
	$\bar{\beta}_0$	$\bar{\delta}$	β	p $/h	p_r $/(100 m)^2	ε_r m	p_t $/(10 min)^2	ε_t min	ω $	P $	Cs$_\pi$ $	Ps$_\pi$ $	Ss$_\pi$ $
1	0.911	0.053	72.22%	0.61	0.048	38.78	0.015	1.00	3.58	0.30	44.37	1.96	46.33
2	0.734	0.048	54.48%	1.14	0.064	51.33	0.036	2.46	3.97	0.56	28.05	2.75	30.80

The higher price leads to peaks of the producer surplus (i.e., profit), indicating for the selected two locations, noon is the most profitable time for the management agency.

It is worth pointing out that given the learning results of the nonlinear regression (21), the time for obtaining a near optimal solution for the large-scale problem is negligible, which demonstrates the efficiency of the proposed algorithm.

6. Conclusion and future work

For short-period parking problems, reservation-based parking systems have the merit of eliminating vehicles cruising for parking, while facing the risk of service failure when users are subject to random late departures. This paper addresses such uncertainty issues and establish a parking reservation system equipped with flexibility. The system could relocate customers or ask customers to wait for an amount of time, where the relocation distance and waiting time are bounded by region flexibility and time flexibility, respectively, which are granted to the system by customers in exchange of price discounts. Given the high nonconvexity and the curse of dimensionality of the problem, a CA based framework is proposed to efficiently obtain a near optimal solution. It is shown that the proposed flexible parking reservation system enables high utilization of parking resources while guaranteeing the service level under customer's departing uncertainty, which is an impossible mission for its non-flexible counterpart. A following case study of a selected downtown area of San Francisco provides several further results. First, it illustrates the efficiency of applying the proposed solution framework to large-scale problems. Second, effects of the spatial heterogeneity on the optimal system equilibrium are discussed, with the key conclusions that pricing decisions and the corresponding flexibility are mainly determined by the lot density, while occupancy at each lot mainly depends on the lot demand pattern, and the unit-area demand function serves as a good indicator for surpluses. Third, with time and region flexibility, the system can better hedge the spatial heterogeneity in parking resources and absorb local demand surges. Finally, we separately compare the proposed system with dynamic pricing based SFpark project and the non-flexible parking reservation system, showing the dominance of the proposed system. With the eliminated amount of cruising traffic, high utilization of parking resources, guaranteed service level, and reduced customer disutility, the proposed parking system promises to solve the parking issues in busy city areas.

Our work can be extended in the following directions. First, we assume all short-term reservations share the same length, and customers parking in the same location share the same disutility, although in reality customers have different parking lengths and disutility. To capture customer heterogeneity, our model can be extended to a version incorporating multi-class demands, where each class of customers shares a fixed parking length and disutility preference. Second, our model only captures the short-term equilibrium where individual drivers' parking needs (where to park and when to park) will not shift based on the parking price. This can be extended through an analogy of dynamic traffic equilibrium model, where demands can be spatial-temporal correlated. Third, the penalties associated with late departures can be endogenized, so that the system can holistically regulate customers' late departure behavior. This may require sufficient data to evaluate the impact of penalty cost on the departure behavior of customers. To increase the practical value of our work, future efforts may include varying the forms of the flexibility disutility function in the numerical study.

Acknowledgments

This research is supported in part by the U.S. National Science Foundation through grants CNS# 1637772. The authors gratefully acknowledge the valuable comments from the editor and four anonymous reviewers that helped improve the paper.

Appendix A. Notation list

(Tables 9 and 10)

Table 9
Notation list of variables and parameters.

Variables		First appearance
p	base price	Page 5, Paragraph 2, Line 8
p_t	price discount per unit time flexibility	Page 5, Paragraph 2, Line 6
p_r	price discount per unit region flexibility	Page 5, Paragraph 2, Line 5
β	reservation occupancy	Page 5, Paragraph 3, Line 2
ε_t	time flexibility	Page 5, Paragraph 1, Line 4
ε_r	region flexibility	Page 5, Paragraph 1, Line 1
$\mathbf{a}$	a sample of reservation given $\boldsymbol{\beta}$	Page 5, Paragraph 4, Line 4
x_{ijs}	binary assignment decision, indexed by customer i, lot j, and stage s	Page 5, Paragraph 6, Line 3
$\mathbf{x}$	$\{x_{ijs}\}$	Page 6, Paragraph 1, Line 2
$\boldsymbol{y}_s$	state variable at stage s	Page 5, Paragraph 6, Line 4
$\mathbf{y}$	$\{\boldsymbol{y}_s\}$	Page 6, Paragraph 1, Line 2
λ_γ	scaled Lagrangian multiplier for the service level constraint	Page 10, Paragraph 1, Line 2
Parameters		
l	length of each reservation	Page 4, Paragraph 6, Line 4
$\bar{l}$	average parking length	Page 5, Paragraph 1, Line 11
n	number of slot at each lot during each planning horizon	Page 4, Paragraph 6, Line 5
β_0	occupancy potential	Page 5, Paragraph 3, Line 7
δ	unit occupancy reduction	Page 5, Paragraph 3, Line 7
τ	length of the decision cycle	Page 5, Paragraph 6, Line 2
$\bar{q}$	service level threshold	Page 7, Paragraph 1, Line 4
ξ^s	realization of the departure uncertainty at the end of stage s	Page 5, Paragraph 6, Line 5
$\xi^{[1,s]}$	$\left\{\xi^1, \xi^2, \ldots, \xi^s\right\}$	Page 6, Paragraph 1, Line 1
ρ	parking lot (single capacity) density	Page 7, Paragraph 3, Line 2
a_t	coefficient for time flexibility disutility	Page 10, Paragraph 4, Line 1
a_r	coefficient for region flexibility disutility	Page 10, Paragraph 4, Line 1
$w, \mathbf{w}_\chi^\intercal, b_\chi, b$	parameters in the nonlinear regression	Page 8, Paragraph 10, Line 2
μ	penalty corresponding to each unserved customer	Page 11, Paragraph 3, Line 5
t_i^r	reservation beginning stage of customer i	Page 11, Paragraph 6, Line 4
t_i^d	reservation end stage of customer i	Page 11, Paragraph 6, Line 4
d_i	reservation destination of customer i	Page 11, Paragraph 6, Line 4
ϵ	termination threshold in determining μ	Page 12, Paragraph 3, Line 6
e	demand elasticity	Page 13, Paragraph 3, Line 9

Table 10
Notation list of functions and other parameters.

Functions		First appearance
D_t	disutility corresponding to the time flexibility	Page 5, Paragraph 2, Line 4
D_r	disutility corresponding to the region flexibility	Page 5, Paragraph 2, Line 4
ω	customer's disutility	Page 5, Paragraph 2, Line 7
Ps	producer surplus	Page 6, Paragraph 3, Line 3
Cs	consumer surplus	Page 6, Paragraph 3, Line 5
Ss	social surplus	Page 6, Paragraph 3, Line 2
q	long-term service failure probability under the best assignment policy	Page 6, Paragraph 1, Line 10
q_π	long-term service failure probability under policy π	Page 8, Paragraph 9, Line 8
Pe	penalty for violating the service constraint	Page 11, Paragraph 3, Line 4
Ss_π	social objective under the assignment policy π	Page 11, Paragraph 1, Line 3
$\mathrm{Ss_u}$	social objective in the relaxed problem obtained under perfect information	Page 11, Paragraph 4, Line 1
$\hat{\mathrm{Ss}}_\mathrm{u}$	social objective in the further relaxed problem	Page 13, Paragraph 1, Line 6
Others		
$\mathcal{J}$	set of parking lots	Page 4, Paragraph 6, Line 6
$\mathcal{T}$	studied rolling horizon	Page 4, Paragraph 6, Line 10
$\mathcal{S}$	set of stages for assignment decisions	Page 5, Paragraph 6, Line 3
$\mathcal{I}$	set of customers	Page 11, Paragraph 6, Line 1
π	defined heuristic assignment policy	Page 8, Paragraph 3, Line 3

Appendix B. Proof of Proposition 1

Proof. Note that in the following context, for the formulation clarity, we abuse e to denote the base of the natural logarithm rather than the demand elasticity.

First, we analyze the scenario of $\varepsilon_r \geq 1$. Denote q_k as the number of customers in the system immediately prior to the beginning of k-th slot, $k \geq 1$. By assuming $\varepsilon_t < l$, we know $q_k \in \{0, 1, 2\}$. With a bit abuse of notation, we define P as the transition probability matrix between q_k and q_{k+1}, i.e., $p_{ij} = P\{q_{k+1} = j | q_k = i\}$. It can be shown that P is

$$\begin{aligned}
p_{00} &= P_{a=0} + P_{a=1}\left(1 - e^{-l/\bar{l}}\right) + P_{a=2}\,p'_{00},\\
p_{01} &= P_{a=1}e^{-l/\bar{l}} + P_{a=2}\,p'_{01},\\
p_{02} &= P_{a=2}\,p'_{02},\\
p_{10} &= P_{a=0}\left(1 - e^{-l/\bar{l}}\right) + P_{a=1}\left(1 - e^{-l/\bar{l}}\right)^2 + P_{a=2}\,p'_{10},\\
p_{11} &= P_{a=0}e^{-l/\bar{l}} + P_{a=1}\binom{2}{1}\left(1 - e^{-l/\bar{l}}\right)e^{-l/\bar{l}} + P_{a=2}\,p'_{11},\\
p_{12} &= P_{a=1}e^{-2l/\bar{l}} + P_{a=2}\,p'_{12},\\
p_{20} &= P_{a=0}\left(1 - e^{-l/\bar{l}}\right)^2 + P_{a=1}\,p'_{10} + P_{a=2}\,p'_{20},\\
p_{21} &= P_{a=0}\binom{2}{1}\left(1 - e^{-l/\bar{l}}\right)e^{-l/\bar{l}} + P_{a=1}\,p'_{11} + P_{a=2}\,p'_{21},\\
p_{22} &= P_{a=0}e^{-2l/\bar{l}} + P_{a=1}\,p'_{12} + P_{a=2}\,p'_{22},
\end{aligned}$$

where

$$\begin{aligned}
p'_{00} &= \left(1 - e^{-\varepsilon_t/\bar{l}}\right)^2\\
p'_{01} &= 2\left(1 - e^{-\varepsilon_t/\bar{l}}\right)e^{-\varepsilon_t/\bar{l}}\\
p'_{02} &= e^{-2\varepsilon_t/\bar{l}},
\end{aligned}$$

$$\begin{aligned}
p'_{10} &= \binom{2}{1}\int_0^{\varepsilon_t} \bar{l}^{-1}e^{-\tau/\bar{l}}\left(1 - e^{-(l-\tau)/\bar{l}}\right)\left(e^{-\tau/\bar{l}} - e^{-l/\bar{l}}\right)d\tau + \left(e^{-\varepsilon_t/\bar{l}} - e^{-l/\bar{l}}\right)^2\\
&= 1 - 4e^{-l/\bar{l}} + e^{-2l/\bar{l}} + 2e^{-(l+\varepsilon_t)/\bar{l}} + 2\varepsilon_t l^{-1}e^{-2l/\bar{l}},
\end{aligned}$$

$$\begin{aligned}
p'_{11} &= \binom{2}{1}e^{-l/\bar{l}}\int_0^{\varepsilon_t} \bar{l}^{-1}e^{-\tau/\bar{l}}\left(1 - e^{-(l-\tau)/\bar{l}}\right)d\tau + \binom{2}{1}\int_0^{\varepsilon_t} \bar{l}^{-1}e^{-\tau/\bar{l}}e^{-(l-\tau)/\bar{l}}\left(e^{-\tau/\bar{l}} - e^{-l/\bar{l}}\right)d\tau + \binom{2}{1}e^{-l/\bar{l}}\left(e^{-\varepsilon_t/\bar{l}} - e^{-l/\bar{l}}\right)\\
&= 4e^{-l/\bar{l}} - 2e^{-2l/\bar{l}} - 2e^{-(l+\varepsilon_t)/\bar{l}} - 4\varepsilon_t\bar{l}^{-1}e^{-2l/\bar{l}},
\end{aligned}$$

$$\begin{aligned}
p'_{12} &= \binom{2}{1}e^{-l/\bar{l}}\int_0^{\varepsilon_t} \bar{l}^{-1}e^{-\tau/\bar{l}}e^{-(l-\tau)/\bar{l}}d\tau + e^{-2l/\bar{l}}\\
&= \left(1 + 2\varepsilon_t\bar{l}^{-1}\right)e^{-2l/\bar{l}},
\end{aligned}$$

$$\begin{aligned}
p'_{20} &= \left(\int_0^{\varepsilon_t} \bar{l}^{-1}e^{-\tau/\bar{l}}\left(1 - e^{-(l-\tau)/\bar{l}}\right)d\tau\right)^2 + \binom{2}{1}\left(e^{-\varepsilon_t/\bar{l}} - e^{-l/\bar{l}}\right)\int_0^{\varepsilon_t} \bar{l}^{-1}e^{-\tau/\bar{l}}\left(1 - e^{-(l-\tau)/\bar{l}}\right)d\tau + \left(e^{-\varepsilon_t/\bar{l}} - e^{-l/\bar{l}}\right)^2\\
&= (1 - e^{-\varepsilon_t/\bar{l}} - \varepsilon_t\bar{l}^{-1}e^{-l/\bar{l}})^2 + 2(e^{-\varepsilon_t/\bar{l}} - e^{-l/\bar{l}})(1 - e^{-\varepsilon_t/\bar{l}} - \varepsilon_t\bar{l}^{-1}e^{-l/\bar{l}}) + (e^{-\varepsilon_t/\bar{l}} - e^{-l/\bar{l}})^2,
\end{aligned}$$

$$\begin{aligned}
p'_{22} &= \left(\int_0^{\varepsilon_t} \bar{l}^{-1}e^{-\tau/\bar{l}}e^{-(l-\tau)/\bar{l}}d\tau\right)^2 + \binom{2}{1}e^{-l/\bar{l}}\int_0^{\varepsilon_t} \bar{l}^{-1}e^{-\tau/\bar{l}}\int_0^{\varepsilon_t-\tau} \bar{l}^{-1}e^{-u/\bar{l}}e^{-(l-\tau-u)/\bar{l}}du\,d\tau\\
&\quad + \binom{2}{1}e^{-l/\bar{l}}\int_0^{\varepsilon_t} \bar{l}^{-1}e^{-\tau/\bar{l}}e^{-(l-\tau)/\bar{l}}d\tau + e^{-2l/\bar{l}}\\
&= \left(1 + 2\varepsilon_t\bar{l}^{-1} + 2\varepsilon_t^2\bar{l}^{-2}\right)e^{-2l/\bar{l}},
\end{aligned}$$

and

$$p'_{21} = 1 - p_{20} - p_{22}.$$

Define $r_m = \lim_{k\to\infty} P\{q_k = m\}$ as the steady state probability of having m customers departing late at the beginning of the k-th slot. The steady state distribution satisfies the following conditions,

$$\mathbf{r} = \mathbf{r}P,$$

$$\sum_{m=0}^{2} r_m = 1,$$

by solving which we obtain

$$r_1 = \left(\frac{p_{11}p_{20} - p_{10}p_{21} - p_{20}}{p_{00}p_{21} - p_{01}p_{20} - p_{21}} + 1 + \frac{p_{11}p_{02} - p_{12}p_{01} - p_{02}}{p_{22}p_{01} - p_{21}p_{02} - p_{01}}\right)^{-1}$$

$$r_2 = \frac{p_{11}p_{02} - p_{12}p_{01} - p_{02}}{p_{22}p_{01} - p_{21}p_{02} - p_{01}} r_1$$

following which

$$\begin{aligned} q &= r_1\beta\left(e^{-\varepsilon_t/\bar{l}}\right)^2/2 + r_2\left((1-\beta)\left(e^{-\varepsilon_t/\bar{l}}\right)^2 + \beta\left(\left(e^{-\varepsilon_t/\bar{l}}\right)^2 + \binom{2}{1}2^{-1}e^{-\varepsilon_t/\bar{l}}\int_0^{\varepsilon_t}\bar{l}^{-1}e^{-\tau/\bar{l}}e^{-(\varepsilon_t-\tau)/\bar{l}}d\tau\right)\right) \\ &= r_1\beta e^{-2\varepsilon_t/\bar{l}}/2 + r_2\left((1-\beta)e^{-2\varepsilon_t/\bar{l}} + \beta\left(e^{-2\varepsilon_t/\bar{l}} + e^{-\varepsilon_t/\bar{l}}\varepsilon_t\bar{l}^{-1}e^{-\varepsilon_t/\bar{l}}\right)\right) \\ &= \left(r_1\beta/2 + r_2\beta\varepsilon_t\bar{l}^{-1} + r_2\right)e^{-2\varepsilon_t\bar{l}^{-1}}. \end{aligned}$$

Then we analyze the scenario of $\varepsilon_r < 1$. When $\varepsilon_r < 1$, the problem reduces to a single lot problem. Therefore, we can focus on only one lot, denoted as lot 1. Denote q_k as the number of customers departing late at the beginning of the k-th slot, $k \geq 1$. By assuming $\varepsilon_t < l$, we know $q_k \in \{0, 1\}$. Define P' as the transition probability matrix between q_k and q_{k+1}, i.e., $p_{ij} = P\{q_{k+1} = j | q_k = i\}$. P is calculated as follows:

$$\begin{aligned} p_{00} &= 1 - \beta + \beta\left(1 - e^{-l/\bar{l}}\right), \\ p_{01} &= \beta e^{-l/\bar{l}}, \\ p_{10} &= (1-\beta)\left(1 - e^{-l/\bar{l}}\right) + \beta\left(e^{-\varepsilon_t/\bar{l}} - e^{-l/\bar{l}} + \int_0^{\varepsilon_t}\bar{l}^{-1}e^{-\tau/\bar{l}}\left(1 - e^{-(l-\tau)/\bar{l}}\right)d\tau\right) \\ &= 1 - e^{-l/\bar{l}} - \bar{l}^{-1}\beta e^{-l/\bar{l}}\varepsilon_t, \\ p_{11} &= (1-\beta)e^{-l/\bar{l}} + \beta\left(e^{-l/\bar{l}} + \int_0^{\varepsilon_t}\bar{l}^{-1}e^{-\tau/\bar{l}}e^{-(l-\tau)/\bar{l}}d\tau\right) \\ &= e^{-l/\bar{l}} + \bar{l}^{-1}\beta e^{-l/\bar{l}}\varepsilon_t. \end{aligned}$$

Similarly we obtain

$$r_0 = \frac{\left(e^{l/\bar{l}} - 1\right)\bar{l} - \beta\varepsilon_t}{\left(\bar{l} - \varepsilon_t\right)\beta + \left(e^{l/\bar{l}} - 1\right)\bar{l}},$$

$$r_1 = \frac{\bar{l}\beta}{\left(\bar{l} - \varepsilon_t\right)\beta + \left(e^{l/\bar{l}} - 1\right)\bar{l}}.$$

Therefore

$$q = r_1 e^{-\varepsilon_t/\bar{l}} = \frac{\bar{l}\beta}{(\bar{l} - \varepsilon_t)\beta + (e^{l/\bar{l}} - 1)\bar{l}}e^{-\varepsilon_t/\bar{l}}.$$

□

Appendix C. KKT conditions for the optimal interior point solution of the π−restricted homogeneous SP

KKT conditions for the optimal interior point solution of the π−restricted homogeneous SP are

$$-n\beta D_r'(\varepsilon_r) - w_2\sqrt{\rho}\lambda_\gamma = 0, \tag{C.1}$$

$$-n\beta D_t'(\varepsilon_t) - w_1 l^{-1}\lambda_\gamma = 0, \tag{C.2}$$

$$n\left(p-\varepsilon_r D_r'(\varepsilon_r)-\varepsilon_t D_t'(\varepsilon_t)\right)-w_3\lambda_\gamma=0, \tag{C.3}$$

$$n\beta-\lambda_\omega\delta=0, \tag{C.4}$$

$$\beta_0-\delta\left(p-\varepsilon_r D_r'(\varepsilon_r)-\varepsilon_t D_t'(\varepsilon_t)+D_r(\varepsilon_r)+D_t(\varepsilon_t)+c\right)-\beta=0, \tag{C.5}$$

$$\bar{\gamma}-b_\chi-w_1 l^{-1}\varepsilon_t-w_2\sqrt{\rho}\varepsilon_r-w_3\beta=0. \tag{C.6}$$

Appendix D. Proof of Corollary 1

Proof. First we show (i).

From (C.1), (C.2) and (C.6), we obtain

$$\varepsilon_t=\frac{w_1 l^{-1}a_r}{w_1^2 l^{-2}a_r+w_2^2\rho a_t}\left(\bar{\gamma}-b_\chi-w_3\beta\right),$$
$$\varepsilon_r=\frac{w_2\sqrt{\rho}a_t}{w_1^2 l^{-2}a_r+w_2^2\rho a_t}\left(\bar{\gamma}-b_\chi-w_3\beta\right).$$

From (C.1), (C.2) and (C.5) we obtain

$$p=\frac{a_t a_r}{w_1^2 l^{-2}a_r+w_2^2\rho a_t}\left(\bar{\gamma}-b_\chi-w_3\beta\right)^2+\frac{\beta_0-\beta}{\delta}-c.$$

And from (C.1) and (C.4), we obtain, respectively,

$$\lambda_\gamma=\frac{-2na_t a_r}{w_1^2 l^{-2}a_r+w_2^2\rho a_t}\beta\left(\bar{\gamma}-b_\chi-w_3\beta\right),$$
$$\lambda_\omega=\frac{n\beta}{\delta}.$$

Substituting above results into (C.3), we obtain that β solves the following equation,

$$\frac{3w_3^2 a_t a_r}{w_1^2 l^{-2}a_r+w_2^2\rho a_t}\beta^2+\left(\frac{-4w_3 a_t a_r\left(\bar{\gamma}-b_\chi\right)}{w_1^2 l^{-2}a_r+w_2^2\rho a_t}+\frac{1}{\delta}\right)\beta-\frac{\beta_0}{\delta}+c+\frac{a_t a_r\left(\bar{\gamma}-b_\chi\right)^2}{w_1^2 l^{-2}a_r+w_2^2\rho a_t}=0. \tag{D.1}$$

Next we show (ii). Left-side of Eq. (D.1) is a quadratic function of β with positive quadratic coefficient. Denote such function as $F(\beta)$. When $-\frac{\beta_0}{\delta}+c+\frac{a_t a_r(\bar{\gamma}-b_\chi)^2}{w_1^2 l^{-2}a_r+w_2^2\rho a_t}<0$ holds, i.e., $F(0)<0$, Eq. (D.1) has a unique positive solution. The condition $\frac{3w_3^2 a_t a_r}{w_1^2 l^{-2}a_r+w_2^2\rho a_t}+\frac{-4w_3 a_t a_r(\bar{\gamma}-b_\chi)}{w_1^2 l^{-2}a_r+w_2^2\rho a_t}+\frac{1}{\delta}-\frac{\beta_0}{\delta}+c+\frac{a_t a_r(\bar{\gamma}-b_\chi)^2}{w_1^2 l^{-2}a_r+w_2^2\rho a_t}>0$, i.e., $F(1)>0$, indicates that the unique solution β^* lies in (0,1).

For part (iii). The objective (13) can be transformed as a cubic function of β with the coefficient of cubic term being negative. Thus, there are at most two β values that satisfy conditions in (i), and the bigger one, if feasible, is the stationary point corresponding to a local maximum. The result in part (ii) indicates that β^* is the local optimum. Since the smaller stationary point is infeasible, β^* is also the global optimum. When boundary conditions for all decision variables are satisfied, the obtained solution is the unique optimal interior point solution. □

Appendix E. Tuning on the relaxed SP

E.1. Piecewise McCormick relaxation for disutility functions

Proposed in the work of Bergamini et al. (2005), piecewise McCormick envelope (with univariate partitioning) strengthens the standard McCormick envelope McCormick (1976) by uniformly partitioning the domain of one variable involved in the bilinear terms into disjoint regions and then constructing McCormick envelopes in each disjoint region. Readers are referred to Castro (2015) for a summary of such method. We apply such method on the two quadratic disutility functions. Define variables $\nu_r^j=\left(\varepsilon_r^j\right)^2$ and $\nu_t^j=\left(\varepsilon_t^j\right)^2$ for each lot $j\in\mathcal{J}$ and denote $\mathbf{v_r}=\left\{\varepsilon_r^j\right\},\mathbf{v_t}=\left\{\varepsilon_t^j\right\}$. Denote N_r as the number of partitions for $\varepsilon_r^j,\forall j\in\mathcal{J}$, and define variables $\varepsilon_{r,m}^j$ and binary indicators $I_{r,m}^j$ for each region $m=1,2,\dots,N_r$, for each lot $j\in\mathcal{J}$. Let ε_r^L and ε_r^U represent the lower and upper bound for $\varepsilon_r^j,\forall j\in\mathcal{J}$, respectively and further denote $\varepsilon_{r,m}^L$ and $\varepsilon_{r,m}^U$, $m=1,2,\dots,N_r$ as the lower and upper bound of the m-th region, respectively. Accordingly, define $\left\{\varepsilon_{t,m}^j\right\}$ and $\left\{I_{t,m}^j\right\}$ and denote N_t, ε_t^L, ε_t^U, $\{\varepsilon_{t,m}^L\}$

and $\{\varepsilon_{t,m}^U\}$ for the time dimension. Based on that, we can relax $\left(\varepsilon_r^j\right)^2$ and $\left(\varepsilon_t^j\right)^2$, $\forall j \in \mathcal{J}$ by the following piecewise McCormick formulations,

$$\left\{\begin{array}{c} v_k^j = \left(\varepsilon_k^j\right)^2, \\ v_k^j \ge \sum_{m=1}^{N_k}\left(2\varepsilon_{k,m}^L\varepsilon_{k,m}^j - \left(\varepsilon_{k,m}^L\right)^2 I_{k,m}^j\right), \\ v_k^j \ge \sum_{m=1}^{N_k}\left(2\varepsilon_{k,m}^U\varepsilon_{k,m}^j - \left(\varepsilon_{k,m}^U\right)^2 I_{k,m}^j\right), \\ v_k^j \le \sum_{m=1}^{N_k}\left(\left(\varepsilon_{k,m}^L + \varepsilon_{k,m}^U\right)\varepsilon_{k,m}^j - \varepsilon_{k,m}^L\varepsilon_{k,m}^U I_{k,m}^j\right), \\ \varepsilon_k^j = \sum_{m=1}^{N_k}\varepsilon_{k,m}^j, \\ \sum_{m=1}^{N_k} I_{k,m}^j = 1, \\ \varepsilon_{k,m}^L = \varepsilon_k^L + \frac{(\varepsilon_k^U - \varepsilon_k^L)(m-1)}{N_k},\ m = 1,2,\ldots,N_k, \\ \varepsilon_{k,m}^U = \varepsilon_k^L + \frac{(\varepsilon_k^U - \varepsilon_k^L)m}{N_k},\ m = 1,2,\ldots,N_k, \\ I_{k,m}^j\varepsilon_{k,m}^L \le \varepsilon_{k,m}^j \le I_{k,m}^j\varepsilon_{k,m}^U,\ m = 1,2,\ldots,N_k, \\ \varepsilon_k^j \in \left[\varepsilon_k^L, \varepsilon_k^U\right], \\ I_{k,m}^j \in \{0,1\},\ m = 1,2,\ldots,N_k, \end{array}\right\}, \quad \forall k \in \{r,t\}, \forall j \in \mathcal{J}. \tag{E.1}$$

We replace $\left(\varepsilon_r^j\right)^2$ and $\left(\varepsilon_t^j\right)^2$in SSP with v_r^j and v_t^j, respectively, for all $j \in \mathcal{J}$ and gain the reformulation of Constraints (10) as

$$\beta^j = \beta_0^j - \delta^j\left(p^j - a_r v_r^j - a_t v_t^j + c\right),\ \ \forall j \in \mathcal{J}. \tag{E.2}$$

Therefore, all constraints in the relaxed SSP are linear. Then we substitute (E.2) into the objective and turn it into a second-order polynomial function of p^j, v_r^j and v_t^j (with linear terms of $\{x_{ijs}\}$) as follows:

$$\mathrm{Ss}_u' = \sum_{j\in\mathcal{J}}\left\{\frac{n\delta}{2}\left[p^j \quad v_r^j \quad v_t^j\right]A_0\left[p^j \quad v_r^j \quad v_t^j\right]^T + B_0^T\left[p^j \quad v_r^j \quad v_t^j\right]^T + C_0\right\} + \mu\sum_{i\in\mathcal{I}_{\mathbf{a}}, j\in\mathcal{J}, s\in\mathcal{S}} x_{ijs}, \tag{E.3}$$

where A_0, B_0 and C_0 are constant coefficients. This sheds some light on efficiently solving the problem since extensive studies have been conducted on solving problems with a quadratic objective if it is semi-definite (refer to Semidefinite Programming (SDP)). Unfortunately, it can be verified that $A_0 = \begin{bmatrix} -1 & 2a_r & 2a_t \\ 2a_r & -3a_r^2 & -3a_r a_t \\ 2a_t & -3a_r a_t & -3a_t^2 \end{bmatrix}$ is not negative semidefinite. Therefore, we further relax the objective into a negative semidefinite function via D.C. decomposition and relaxation.

E.2. D.C. decomposition and relaxation

In this section, we will construct a concave relaxation of the relaxed SSP objective (E.3) by overestimating the convex terms in the equivalent separable form of (E.3) after eigen-transformation. The resulting relaxed objective is concave and therefore is manageable with existing methods in SDP studies. Following a special D.C. decomposition scheme with diagonal perturbation on A_0 (Bomze, 2002; Zheng et al., 2011), the quadratic term in the objective is equivalent to

$$\frac{n\delta}{2}\sum_{j\in\mathcal{J}}\left\{\left[p^j \quad v_r^j \quad v_t^j\right]\left[A_0 - \mathrm{Diag}(\sigma)\right]\left[p^j \quad v_r^j \quad v_t^j\right]^T + \left[p^j \quad v_r^j \quad v_t^j\right]\mathrm{Diag}(\sigma)\left[p^j \quad v_r^j \quad v_t^j\right]^T\right\},$$

where $\mathrm{Diag}(\sigma)$ is the diagonal matrix with σ_i being the i-th diagonal element. By taking $\sigma = |\lambda_{\max}(A_0)|\,\boldsymbol{e}$, where $\lambda_{\max}(A_0) = \left(\sqrt{4(a_r^2+a_t^2)+(1+3a_r^2+3a_t^2)^2} - (1+3a_r^2+3a_t^2)\right)/2$ is the maximum eigenvalue of A_0 and $\boldsymbol{e} = (1,\ldots,1)^T$, $[A_0 - \mathrm{Diag}(\sigma)] \preceq 0$ is guaranteed. By setting realistic bounds $p^j \in [0, 2p']$, $\varepsilon_r^j \in \left[0, \frac{l}{2}\right]$ and $\varepsilon_r^j \in [0, \max(d)]$ for all $j \in \mathcal{J}$, where $\max(d)$ is the maximum distance between two lots in the given example (Node that in a bigger problem instance, ε_r^j could be set as a realistic constant value.), we can overestimate $\sum_{j\in\mathcal{J}}\left\{\left[p^j \quad v_r^j \quad v_t^j\right]\mathrm{Diag}(\sigma)\left[p^j \quad v_r^j \quad v_t^j\right]^T\right\}$ with the following linear function

$$\sum_{j\in\mathcal{J}}\left\{2\sigma p' p^j + \frac{\sigma l^2}{4}v_r^j + \sigma(\max(d))^2 v_t^j\right\}.$$

Therefore, the relaxed SSP objective (E.3) can be overestimated by

$$\frac{n\delta}{2}\sum_{j\in\mathcal{J}}\Big\{\big[p^j \quad v_r^j \quad v_t^j\big][A_0-\mathrm{Diag}(\sigma)]\big[p^j \quad v_r^j \quad v_t^j\big]^T+2\sigma p'p^j+\frac{\sigma l^2}{4}v_r^j+\sigma(\max(d))^2v_r^j$$
$$+B_0^T\big[p^j \quad v_r^j \quad v_t^j\big]^T+C_0\Big\}+\mu\sum_{i\in\mathcal{I}_\mathbf{a},j\in\mathcal{J},s\in\mathcal{S}}x_{ijs}, \tag{E.4}$$

which, by construction, is a concave quadratic function.

In the following we present the relaxed SP for a single uncertainty scenario with perfect information.

$$\max_{\mathbf{p},\mathbf{p_r},\mathbf{p_t},\boldsymbol{\varepsilon_r},\boldsymbol{\varepsilon_t},\mathbf{v_r},\mathbf{v_t},\boldsymbol{\beta},\mathbf{x},\mathbf{y}} \hat{Ss}_u=(E.4)$$
$$\text{s.t.}(29),(E.2),(12),$$
$$(E.1),$$
$$(32)-(39).$$

References

Anderson, S.P., De Palma, A., 2004. The economics of pricing parking. J. Urban Econ. 55 (1), 1–20.
Ansari, S., Başdere, M., Li, X., Ouyang, Y., Smilowitz, K., 2017. Advancements in continuous approximation models for logistics and transportation systems: 1996–2016. Transp. Res. Part B.
Arnott, R., 2014. On the optimal target curbside parking occupancy rate. Econ. Transp. 3 (2), 133–144.
Arnott, R., De Palma, A., Lindsey, R., 1991. A temporal and spatial equilibrium analysis of commuter parking. J. Public Econ. 45 (3), 301–335.
Arnott, R., Inci, E., Rowse, J., 2015. Downtown curbside parking capacity. J. Urban Econ. 86, 83–97.
Aurenhammer, F., 1991. Voronoi diagrams - a survey of a fundamental geometric data structure. ACM Comput. Surv. (CSUR) 23 (3), 345–405.
Ayala, D., Wolfson, O., Xu, B., DasGupta, B., Lin, J., 2012. Pricing of parking for congestion reduction. In: Proceedings of the 20th International Conference on Advances in Geographic Information Systems. ACM, pp. 43–51.
Bai, Y., Li, X., Peng, F., Wang, X., Ouyang, Y., 2015. Effects of disruption risks on biorefinery location design. Energies 8 (2), 1468–1486.
Bai, Y., Ouyang, Y., Pang, J.-S., 2016. Enhanced models and improved solution for competitive biofuel supply chain design under land use constraints. Eur. J. Oper. Res. 249 (1), 281–297.
Bergamini, M.L., Aguirre, P., Grossmann, I., 2005. Logic-based outer approximation for globally optimal synthesis of process networks. Comput. Chem. Eng. 29 (9), 1914–1933.
Bomze, I.M., 2002. Branch-and-bound approaches to standard quadratic optimization problems. J. Global Optim. 22 (1–4), 17–37.
Bookbinder, J.H., Desilets, A., 1992. Transfer optimization in a transit network. Transp. Sci. 26 (2), 106–118.
Brown, D.B., Smith, J.E., Sun, P., 2010. Information relaxations and duality in stochastic dynamic programs. Oper. Res. 58 (4-part-1), 785–801.
Buckingham, E., 1914. On physically similar systems; illustrations of the use of dimensional equations. Phys. Rev. 4 (4), 345.
Castro, P.M., 2015. Tightening piecewise McCormick relaxations for bilinear problems. Comput. Chem. Eng. 72, 300–311.
Chen, Z., Yin, Y., He, F., Lin, J.L., 2015. Parking reservation for managing downtown curbside parking. Transp. Res. Rec. (2498) 12–18.
Chiou, S.-W., 2005. Bilevel programming for the continuous transport network design problem. Transp. Res. Part B 39 (4), 361–383.
Fisher, M.L., 1981. The lagrangian relaxation method for solving integer programming problems. Manag Sci. 27 (1), 1–18.
Garey, M.R., Johnson, D.S., 2002. Computers and Intractability, 29. wh freeman New York.
He, F., Yin, Y., Chen, Z., Zhou, J., 2015. Pricing of parking games with atomic players. Transp. Res. Part B 73, 1–12.
Hoogendoorn, S.P., Bovy, P.H., 2004. Pedestrian route-choice and activity scheduling theory and models. Transp. Res. Part B 38 (2), 169–190.
Idris, M., Leng, Y., Tamil, E., Noor, N., Razak, Z., 2009. Car parking system: a review of smart parking system and its technology. Inf. Technol. J. (2).
Jahn, O., Möhring, R.H., Schulz, A.S., Stier-Moses, N.E., 2005. System-optimal routing of traffic flows with user constraints in networks with congestion. Oper. Res. 53 (4), 600–616.
Kuhn, H.W., Tucker, A.W., 1951. Nonlinear programming. In: Proceedings of 2nd Berkeley Symposium. Berkeley: University of California Press, pp. 481–492.
Langevin, A., Mbaraga, P., Campbell, J.F., 1996. Continuous approximation models in freight distribution: an overview. Transp. Res. Part B 30 (3), 163–188.
Lei, C., Ouyang, Y., 2017. Dynamic pricing and reservation for intelligent urban parking management. Transp. Res. Part C 77, 226–244.
Lim, M.K., Mak, H.-Y., Shen, Z.-J.M., 2016. Agility and proximity considerations in supply chain design. Manag. Sci. 63 (4), 1026–1041.
Liu, W., Yang, H., Yin, Y., 2014. Expirable parking reservations for managing morning commute with parking space constraints. Transp. Res. Part C 44, 185–201.
Mackowski, D., Bai, Y., Ouyang, Y., 2015. Parking space management via dynamic performance-based pricing. Transp. Res. Part C 59, 66–91.
McCormick, G.P., 1976. Computability of global solutions to factorable nonconvex programs: part i: convex underestimating problems. Math. Program. 10 (1), 147–175.
Newell, G.F., 1971. Dispatching policies for a transportation route. Tranp. Sci. 5 (1), 91–105.
NYC-OpenData, 2017. Parking violations issued - fiscal year 2017. https://data.cityofnewyork.us/City-Government/Parking-Violations-Issued-Fiscal-Year-2017/2bnn-yakx. Accessed: 2018-12-10.
Ouyang, Y., Daganzo, C.F., 2006. Discretization and validation of the continuum approximation scheme for terminal system design. Tranp. Sci. 40 (1), 89–98.
Ouyang, Y., Nourbakhsh, S.M., Cassidy, M.J., 2014. Continuum approximation approach to bus network design under spatially heterogeneous demand. Transp. Res. Part B 68, 333–344.
Parzen, E., 1962. On estimation of a probability density function and mode. Ann. Math. Stat. 33 (3), 1065–1076.
Qian, Z., Rajagopal, R., 2015. Optimal dynamic pricing for morning commute parking. Transportmetrica A 11 (4), 291–316.
Qian, Z.S., Rajagopal, R., 2014. Optimal occupancy-driven parking pricing under demand uncertainties and traveler heterogeneity: a stochastic control approach. Transp. Res. Part B 67, 144–165.
Qian, Z.S., Xiao, F.E., Zhang, H., 2012. Managing morning commute traffic with parking. Transp. Res. Part B 46 (7), 894–916.
Rockafellar, R.T., Wets, R.J.-B., 1991. Scenarios and policy aggregation in optimization under uncertainty. Mathematics of Oper. Res. 16 (1), 119–147.
SFMTA, 2014a. Pilot project evaluation. http://direct.sfpark.org/wp-content/uploads/eval/SFpark_Pilot_Project_Evaluation.pdf. Accessed: 2018-06-29.
SFMTA, 2014b. Sfpark parking sensor data hourly occupancy 2011 - 2013. https://striping.sfmta.com/sfpark/SFpark_ParkingSensorData_HourlyOccupancy_20112013.csv. Accessed: 2018-10-20.
SFMTA, 2018. Parking meters. https://data.sfgov.org/Transportation/Parking-meters/44tz-8bvs. Accessed: 2018-06-05.
Shao, C., Yang, H., Zhang, Y., Ke, J., 2016. A simple reservation and allocation model of shared parking lots. Transp. Res. Part C 71, 303–312.
Shoup, D., 2005. The High Cost of Free Parking. American Planning Association.
Shoup, D.C., 2006. Cruising for parking. Transp. Policy 13 (6), 479–486.

Smilowitz, K.R., Daganzo, C.F., 2007. Continuum approximation techniques for the design of integrated package distribution systems. Networks 50 (3), 183–196.
TCRP, 2005. Traveler Response to Transportation System Changes Handbook, Third Edition: Chapter 13, Parking Pricing and Fees. Transp. Res. Board.
Wang, X., Lim, M.K., Ouyang, Y., 2016. A continuum approximation approach to the dynamic facility location problem in a growing market. Tranp. Sci. 51 (1), 343–357.
Wang, X., Lim, M.K., Ouyang, Y., 2017. Food-energy-environment trilemma: policy impacts on farmland use and biofuel industry development. Energy Econ. 67, 35–48.
Wang, X., Ouyang, Y., 2013. A continuum approximation approach to competitive facility location design under facility disruption risks. Transp. Res. Part B 50, 90–103.
Xiao, H., Xu, M., Gao, Z., 2018. Shared parking problem: a novel truthful double auction mechanism approach. Transp. Res. Part B 109, 40–69.
Xu, S.X., Cheng, M., Kong, X.T., Yang, H., Huang, G.Q., 2016. Private parking slot sharing. Transp. Res. Part B 93, 596–617.
Yang, H., Bell, M.G., 1997. Traffic restraint, road pricing and network equilibrium. Transp. Res. Part B 31 (4), 303–314.
Yang, H., Liu, W., Wang, X., Zhang, X., 2013. On the morning commute problem with bottleneck congestion and parking space constraints. Transp. Res. Part B 58, 106–118.
Yang, H., Wang, X., 2011. Managing network mobility with tradable credits. Transp. Res. Part B 45 (3), 580–594.
Zhang, X., Yang, H., Huang, H.-J., 2011. Improving travel efficiency by parking permits distribution and trading. Transp. Res. Part B 45 (7), 1018–1034.
Zheng, N., Geroliminis, N., 2016. Modeling and optimization of multimodal urban networks with limited parking and dynamic pricing. Transp. Res. Part B 83, 36–58.
Zheng, X., Sun, X., Li, D., 2011. Nonconvex quadratically constrained quadratic programming: best dc decompositions and their sdp representations. J. Global Optim. 50 (4), 695–712.