UC Irvine
UC Irvine Electronic Theses and Dissertations

Title
Control Theoretic Approaches to Congestion Pricing for High-occupancy Toll Lanes

Permalink
https://escholarship.org/uc/item/0fs4t9kz

Author
Wang, Xuting

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Control Theoretic Approaches to Congestion Pricing for High-occupancy Toll Lanes

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Civil Engineering

by

Xuting Wang

Dissertation Committee:
Professor Wen-Long Jin, Chair
Professor Will Recker
Professor R. Jayakrishnan

2019

DEDICATION

To my family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

CURRICULUM VITAE

Xuting Wang

EDUCATION

Doctor of Philosophy in Civil Engineering   **2019**
University of California, Irvine   *Irvine, California*

Master of Science in Civil Engineering   **2016**
University of California, Irvine   *Irvine, California*

Bachelor of Science in Transportation   **2014**
Southwest Jiaotong University   *Chengdu, Sichuan*

RESEARCH EXPERIENCE

Graduate Research Assistant   **2015–2019**
University of California, Irvine   *Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant   **Spring 2018, Winter 2019**
University of California, Irvine   *Irvine, California*

REFEREED CONFERENCE PUBLICATIONS

Revenue maximization for high-occupancy toll lanes: **Jan 2019**
An optimal control approach
Transportation Research Board 98th Annual Meeting

Control and estimation of traffic systems with high-   **Jan 2018**
occupancy toll lanes
Transportation Research Board 97th Annual Meeting

A new method to estimate value of time for high-   **Jan 2017**
occupancy-toll lane operation
Transportation Research Board 96th Annual Meeting

ABSTRACT OF THE DISSERTATION

Control Theoretic Approaches to Congestion Pricing for High-occupancy Toll Lanes

By

Xuting Wang

Doctor of Philosophy in Civil Engineering

University of California, Irvine, 2019

Professor Wen-Long Jin, Chair

The purpose of this study is to propose control theoretic approaches for high-occupancy toll (HOT) lanes operation. This dissertation considers different operation objectives, and provides pricing schemes for HOT lanes accordingly.

To improve the system performance, the study first proposes a simultaneous estimation and control method for the same system as that in (Yin and Lou, 2009). An integral controller is applied to estimate the average value of time (VOT) of SOVs, and the dynamic prices are calculated based on the logit model. The closed-loop system is proved to be stable and guaranteed to converge to the optimal state both analytically and numerically. Two convergence patterns, Gaussian or exponential, are revealed. The effect of the scale parameter in the logit model is also examined.

Then, a new lane choice model, i.e., the vehicle-based user equilibrium principle, is proposed to capture the lane choice of SOVs. A general lane choice model is derived based on the characteristics of the logit and the vehicle-based UE model. An insight regarding the dynamic price is obtained by analytically solving the optimal dynamic prices with constant demands of HOVs and SOVs, and then a feedback controller is designed to determine the dynamic prices without knowing SOVs' lane choice models, but to satisfy the two control objectives: maximizing the flow-rate but not forming a queue on the HOT lanes. If the type of the lane

choice model is given, the distribution of VOTs of the SOVs can be estimated.

Next, an optimal control problem is proposed to examine the statement that revenue maximization should generally coincide with the optimization of freeway performances, such as maximizing overall travel-time savings or throughput. Results show that operators need to make different strategies based on the traffic demand. In order to maximize the revenue, operators should set a higher price to make the HOT lanes underutilized if the demand of HOVs is low. However, if the demand of HOVs is high, operators need to set a lower price to attract more SOVs to create congestion on the HOT lanes.

It has long been known that drivers' departure time choice behavior is one fundamental cause of congestion. In the last part of this dissertation, pricing schemes are proposed to consider both lane choice and departure time choice. In the study period, the demands for the HOT and GP lanes are higher than their capacities, which means the whole freeway is congested. However, the congestion period on the HOT lanes is short than that on the GP lanes. So, the HOT lanes are "underutilized". It turns out that flat (instead of dynamic) pricing schemes are able to meet the following two constraints: (1) the total travel time and scheduling cost is minimized; and (2) the costs for each non-switching and switching SOV are the same. We show that different revenue and tolling constrains for certain type of vehicles lead to different pricing schemes.

Chapter 1

Introduction

1.1 Research Background

In recent years, traffic congestion has become more severe in large cities and on highways, and it affects economic competitiveness, driving safety, and air quality. Schrank et al. (2015) reported that during peak hours, the average annual delay in 2014 for an auto commuter in Los Angeles was about 80 hours, and the average congestion cost was 1711 dollars. Transportation agencies have tried different strategies to ease congestion, such as expanding road capacity, optimizing signal timing and control, and promoting voluntary reduction in driving at businesses and other large organizations.

In the U.S., high-occupancy vehicle (HOV) lane (also known as carpool lane) are widely used on freeways to reduce congestion. This lane is available for cars with a minimum of two or three occupants. However, some HOV lanes could be underutilized, even when the corresponding general purpose (GP) lanes on the same roads are congested.

Congestion pricing has received more and more attention in both economics and transporta-

tion fields since 1920s. A type of relatively recent congestion pricing strategy is realized with high-occupancy toll (HOT) lanes, where single-occupancy vehicles (SOVs) can pay a price to use HOV lanes during peak periods. The first HOT lane has been implemented on SR-91 in California since 1995. As of January 2019, 41 HOT lanes had been implemented nationwide. So far HOT lanes implemented in the real world have been operated based on heuristic pricing schemes, which are determined based on the historic and current traffic conditions. However, researches on HOT lanes operation are still limited, especially based on the control theory. This dissertation proposes control theoretic approaches to congestion pricing for HOT lanes which attach importance to both guidelines for operation schemes and understanding of characteristics of SOV drivers.

1.2 Research Objectives

The study site of this dissertation is a freeway corridor between an origin and a destination. The freeway has two types of lanes: the HOT and GP lanes. There are two bottlenecks, one for each lane group. This study aims to propose control theoretic approaches that: (1) providing pricing schemes for HOT lanes to achieve various operation objectives; and (2) revealing the characteristics of SOV drivers. To achieve those goals, the detailed objectives are described as follows:

- **To simultaneously estimate the average value of time (VOT) of SOVs and determine dynamic pricing schemes for HOT lanes.**

  Traditionally, except the work by (Yin and Lou, 2009) and their follow-up studies, VOTs have been estimated offline. We apply a simple feedback controller to realize online estimation of the average VOT of SOVs, based on the two operation objectives of HOT lanes, i.e., keeping zero queue on the HOT lanes and maximizing the HOT lanes throughput. Then, we can calculate the dynamic price for the HOT lanes. This framework enables the analytical and numerical proof that the closed-loop system is stable and guaranteed to converge to the optimal state.

- **To design a stable dynamic pricing scheme independent of lane choice models for HOT lanes.**

  With the same operation objectives as those in the first part, we develop a feedback controller to determine the dynamic prices without knowing SOVs' lane choice models. However, if the type of lane choice models is known, we can estimate the VOTs.

- **To develop a pricing scheme to maximize the revenue for HOT lane operators.**

  Some agencies claimed that revenue maximization should generally coincide with the

optimization of freeway performances. But this statement has not been carefully examined. In this dissertation, we apply the optimal control theory to formulate and solve the revenue maximization problem when the freeway is congested.

- **To propose pricing schemes for HOT lanes considering departure time user equilibrium.**

  The departure time choice behavior is one fundamental cause of congestion. In this dissertation, we design pricing schemes for HOT lanes considering commuters' departure time choice. The flat pricing scheme enables future studies on the car sharing problem, as well as the auction-based pricing scheme for HOT lanes.

1.3   Research Outline

The overall dissertation framework is described in Figure 1.1.
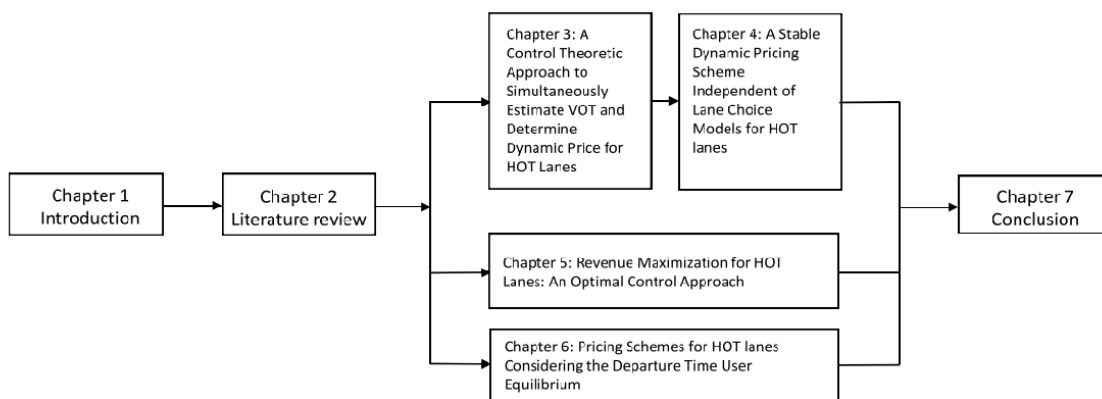
Figure 1.1: Dissertation framework

Chapter 1 introduces the research background and objectives of this dissertation.

Chapter 2 provides a detailed literature review regarding the five important components of this dissertation: congestion pricing, the development of HOT lanes, lane choice models, value of time, and departure time choice.

Chapter 3 makes three fundamental contributions for the theoretic studies on HOT lanes: (1) to present a simpler formulation of the point queue model based on the new concept of residue capacity; (2) to propose a simple feedback control theoretic approach to estimate the average value of time and calculate the dynamic price; and (3) to analytically and numerically prove that the closed-loop system is stable and guaranteed to converge to the optimal state, in either Gaussian or exponential manners. We provide numerical examples to show that our method is effective and robust with respect to variations in the demand pattern. At last, we numerically show that the scale parameter in the logit model does not affect the estimation

of VOT and the optimal state after a long time.

Chapter 4 proposes a general lane choice model based on characteristics of the logit model and the vehicle-based user equilibrium principle. From the general lane choice model, we develop a feedback controller to determine the dynamic prices without knowing SOVs lane choice models, but to satisfy the two control objectives: to maximize the flow-rate but not to form a queue on the HOT lanes. Similar to Chapter 3, we provide equilibrium state and stability analysis for the system. If the type of lane choice models is known, the VOTs can be estimated accordingly. With constant and random demands of HOVs and SOVs, we numerically check the effectiveness of the proposed controller considering: a logit model and a vehicle-based UE principle with negative exponentially distributed VOTs.

Chapter 5 investigates the revenue problem for operating HOT lanes. Some agencies claimed that revenue maximization should generally coincide with the optimization of freeway performances. But this statement has not been carefully examined. In this study, we apply the optimal control theory to formulate and solve the revenue maximization problem when the freeway is congested. We first present the optimality conditions for this optimal control problem, and then provide an iterative method to solve the problem. We compare the revenue and system performance with our previous studies in the same numerical setups. Results show that the performance optimization and revenue maximization cannot be achieved at the same time.

Chapter 6 proposes pricing schemes for HOT lanes considering departure time user equilibrium. From our daily observation, both HOT and GP lanes can be congested during peak periods. The departure time choice behavior has been proved to be one fundamental cause of congestion. We propose three flat pricing schemes that satisfy the following criteria: (1) the total travel time and scheduling cost is minimized; and (2) the costs for each non-paying and paying SOV are the same. We provide departure time equilibrium analysis for the no-toll and toll scenarios, and compare the performance under different pricing schemes. We also

6

compare pricing schemes with and without consideration of departure time choice.

Chapter 7 summarizes the results and findings of this dissertation, and proposes potential research topics.

Chapter 2

Literature Review

This chapter reviews the five elements in the research framework. Section 2.1 reviews the history of congestion pricing in a general network. Section 2.2 reviews the development of HOT lanes and related researches. Section 2.3 reviews two types of choice models that have been applied to capture SOVs' lane choice behavior. Section 2.4 illustrates the the role of VOT in the choice model, and review existing methods for estimating VOT. Section 2.5 reviews the departure time choice models.

2.1 Congestion Pricing

Pigou (1920) first proposed the idea of road pricing. In the paper, it is suggested that an optimal charge should be implemented for the congested road to internalize the externality of vehicles and drive the system to an optimal state. Knight (1924) expressed the same idea by stating that the government should levy a small tax on each truck using the narrow road, so that their travel time cost, plus the tax, is equal to the cost on the broad road. de Palma and Lindsey (2011) pointed that a negative externality is created since travelers

impose delays on others, but they do not pay the full marginal social cost of their trip. In this case, congestion pricing should be imposed to internalize the costs of a negative externality. Sorensen et al. (2008) indicated that many strategies (such as raising fuel prices) usually only provide short-term relieves, but only pricing strategies could manage congestion in the long run.

Congestion pricing strategies can be classified based on different criteria: pricing theory (first-based pricing or second-best pricing), type of analysis (static pricing or dynamic pricing), the location (link-based, path-based, or zone-based pricing), and operation objectives (minimizing total travel delay, or maximizing social welfare, or maximizing revenues for operators).

Static and Dynamic Pricing

Regarding time dimension, there are two strategies in congestion pricing: static and dynamic congestion pricing. Static pricing assumes that congestion is constant over some given time period, the toll typically should stay the same. For example, most vehicles need to pay a fee of £11.50 for entering the Central London between 07:00 and 18:00, Mondays to Fridays. Ieromonachou et al. (2006) indicated that traffic congestion in Oslo, was reduced by 5 percent within the first year of operation after implementing a static toll.

However, the static pricing schemes do not consider the characteristics of traffic dynamics, and it may not work during the peak hour. In this case, dynamic pricing strategies should be implemented. Vickrey (1969) assumed a scenario in which a bottleneck with finite capacity exists on the way for commuters to work during peak hours, and a piecewise linear pricing structure was proposed to eliminate the queue. Later, Vickrey's bottleneck model has been incorporated into more complicated scenarios, such as departure time choice (Hendrickson and Kocur, 1981; Arnott et al., 1987) and more complicated network (Yin and Lou, 2009).

9

First-best and Second-best Pricing

First-best pricing and second-best pricing schemes are widely used for dynamic pricing.

The goal of first-best pricing is to reach a system optimal (SO) flow, the toll rate is equal to the difference between marginal social cost and marginal private cost (Pigouvian tax). Walters (1961) pointed out that the marginal social cost should be equal to the marginal private cost multiplied by elasticity. To be more specific, if $C(n)$ represents the unit private cost when the the number of vehicles on the road link is n per hour, the toll should be set as $n\frac{dC(n)}{dn}$. Dafermos (1972) proposed an algorithm to find the SO flow for a general congested transportation network with multiple vehicle types. Yang and Huang (1998) demonstrated how marginal-cost pricing could be applied in a road network in the presence of elastic demand and queues.

Some assumptions to realize the first-best pricing are: (1) individual drivers choose the route in a rational manner, which is based on utility maximization principle; (2) congestion pricing is applied to all relevant road segments in the network; (3) There is full information on all costs involved (including detours) for both the operators and the drivers; (4) congestion pricing is technically feasible and the transaction costs are reasonably low (Emmerink et al., 1995).

A congestion-pricing scheme is first-best only if all the first-best assumptions are satisfied; otherwise the scheme is second-best. Two key problems related to second-best pricing strategy are: (1) toll rates and locations; and (2) different impacts introduced by toll to heterogeneous drivers (Yang and Huang, 2005). Marchand (1968) extended Levy-Lambert (1968)'s work, and derived a second-best pricing strategy for two-route scenario to maximize a linear combination of the individual's utility. Verhoef et al. (1996a) investigated the optimal one-route toll, and remarked importance of considering the "spill-over" effects on the route without tolls. May and Milne (2000) concluded four practical tolling alternatives: travel-

10

time based charging, travel-distance based charging, link-based charging and cordon-based charging.

2.2 High-occupancy Toll Lanes

High-occupancy vehicle (HOV) lanes are those reserved for cars with a minimum of two or three occupants and other qualified vehicles. The first HOV lane was implemented on Virginia's Shirley Highway busway facility (I-395) in 1969, as shown in Figure 2.1; as of 2005, HOV lanes comprised 1,305 (directional) lane miles of freeway in California, and additional 950 lane miles had been proposed for construction (Jang et al., 2009). HOV lanes can improve the people-moving capability and reliability, encourage car-sharing and reduce congestion, and lead to more efficient usage of the available roadway infrastructure and transit system (Fuhs and Obenberger, 2002). However, in some cases, HOV lanes could be underutilized, even when the parallel GP lanes on the same freeway are congested. Kwon and Varaiya (2008) investigated data from more than 700 detector stations in California during PM peak hours on 128 weekdays in 2005, and found that the flow-rates of 81% HOV lanes were below 1400 vphpl, and most of them had speeds over 45 mph and thus were uncongested and underutilized.

Figure 2.1: The first HOV lane in the U.S., in the Henry G. Shirley Memorial Highway in Northern Virginia

A type of relatively recent congestion pricing strategies are realized with high-occupancy toll (HOT) lanes, where single-occupancy vehicles (SOVs) can pay a price to use HOV lanes during peak periods. The first HOT lane has been implemented on SR-91 in California since 1995; the first HOV-to-HOT conversion project started operation on I-15 near San Diego in 1996; as of May 2012, 14 HOT lanes had been implemented, and additional 14 facilities were under construction (Perez et al., 2012); as of January 2019, 41 HOT lanes had been operated nationwide (TRB Managed Lanes Committee, 2019). As can be seen in Figure 2.2, the number of HOT lanes in the U.S. has increased extremely fast during the last twenty years. In addition to improving the performance of the overall system with better utilization of underutilized HOV lanes, some other benefits from the HOT lanes are: (1) protecting environment by providing opportunities to encourage carpooling, improving transit service

and moving more people in fewer vehicles at faster speeds; and (2) generating new revenue sources that can be used to support the construction of the existing HOT lanes and other initiatives (Perez and Sciara, 2003).
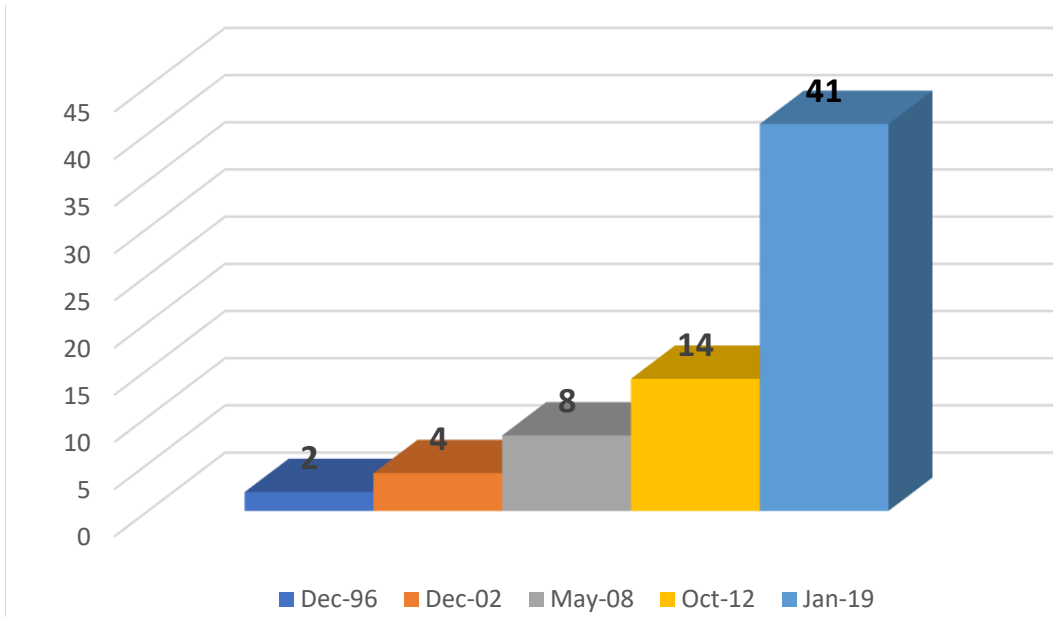
Figure 2.2: Exponential increase in the number of HOT lanes in the U.S.

Typical objectives of operating HOT lanes are: (1) maintaining a certain level of service (LOS) on the HOT lanes, which includes keeping free-flow speed, keeping zero queue, and keeping the speed above certain value; (2) improving the overall system performance (e.g., maximizing vehicle or person throughput, which is equivalent to minimize the travel time); and (3) maximizing revenue for the operators (Perez et al., 2012). Figure 2.3 summarizes the corresponding toll policy, toll rate and operation objectives for 6 well-known HOT lanes in the U.S.. A more detailed overview of the existing 12 HOT facilities in the U.S. can be found in (Chung and Recker, 2011).

Location	Toll Policy		Toll Rate
SR-91	Peak Hour	Off-peak Hour	Time-of-day tolls
Orange County	SOV, HOV2 fully tolled	fully tolled	$1.45 - $9.85
CA	HOV 3+ 50% toll off	free	per use.
I-15 North	SOV tolled		Dynamic tolls
San Diego, CA	HOV2+ free		$0.5 - $8 per use
I-394W	Peak Hour	Off-peak Hour	Dynamic tolls
Minneapolis,	SOV tolled	free	$0.25 - $8,
MN	HOV2+ free	free	change every 3 minutes
I-95	SOV, HOV2 tolled		Dynamic tolls
Miami FL	HOV3+ free(registration required)		$0.25 - $7.25 per use
I-110 &	Peak Hour	Off-peak Hour	Dynamic tolls
I-10	SOV tolled	tolled	by distance
Los Angeles	HOV2 tolled	free	$0.25-$1.4 per mile
CA	HOV3+ free	free	(FasTrak required)

(a) Policy

Location	Project Objective
SR-91	Optimize throughput at FFS [1],
Orange County	generating sufficient revenue for the operations
CA	and maintenance of the toll lanes
I-15 North	Maximizing the use of the existing I-15 ExpressLanes,
San Diego, CA	funding new transit and HOV improvements in this corridor
I-394W	Maintaining FFS for transit and carpools,
Minneapolis,	
MN	using excess revenue to improve transit and highway in I-394 corridor
I-95	Maximizing freeway throughput,
Miami FL	maintaining FFS on the Express Lanes
I-110 &	
I-10	Maximizing efficiency of the entire freeway,
Los Angeles	keep traffic in the ExpressLanes moving smoothly
CA	

(b) Objectives

Figure 2.3: Some typical HOT facilities in the U.S.

So far HOT lanes implemented in the real world have been operated based on heuristic pricing schemes, which are determined based on the historic and current traffic conditions. For example, the original SR-91 Express Lanes in the Orange County had one single entry and one single exit in each direction, the toll collection can be categorized as per use-based. Since March 2017, the Express Lanes have been extended to Riverside County, and drivers may drive the entire length or enter or exit at the county line near the Green River Road; thus the new prices on the extended SR-91 Express Lanes are both time-of-day and distance-based. The prices are adjusted every three months, based on the average hourly flow-rate in the last 12 consecutive weeks (OCTA, 2003). During Phase I of the I-15 Express lanes near San Diego, a limited number of SOVs paid a flat monthly fee to get access to the Express Lanes; during the ongoing Phase II (started in March 1998), the I-15 Express Lanes apply a distance-based dynamic pricing scheme for SOVs between $0.50 and $8.00 to maintain LOS C on the Express lanes, and the price was updated once every 6 minutes initially, or every 3 minutes more recently based on the level of traffic of the Express Lanes (Supernak et al., 2003; Brownstone et al., 2003). However, these heuristic pricing schemes cannot guarantee that the overall traffic system is at the optimal state. In some cases, they would increase the price when the HOT lanes are uncongested and severely underutilized with a very low

flow-rate even if the corresponding GP lanes are congested, cannot eliminate congestion on the HOT lanes after a queue forms on them and the demand is relatively high, and may not be able to cope with accidents on the toll lanes. For example, the SR-91 Express Lanes increase the hourly toll by \$1.00 if the average flow-rate is 3300 vph or more during peak hours (OCTA, 2003). However, from the basic traffic flow theory, the flow-rate itself cannot describe whether the traffic is congested or not. Less SOVs would like to pay and switch if the operators increase price when the Express Lanes are uncongested, thus the congestion on the GP lanes will be even worse.

In the literature, most studies consider the first two operational objectives, which guarantee that the trip time reliability of both HOVs and paying SOVs and help minimize the delay for non-paying SOVs on the GP lanes. But they differ in their estimation of VOTs, pricing strategies as well as the underlying traffic flow and lane choice models. Yin and Lou (2009) proposed a feedback method and a self-learning method to determine the dynamic price to provide a free-flow traffic condition on the HOT lanes while maximizing the throughput of the freeway. For both methods, they used a logit model to capture the lane choice of SOVs in a freeway with HOV, HOT and GP lanes, and a point queue model to capture traffic dynamics. But different from (Brownstone et al., 2003; Lam and Small, 2001), they applied the Kalman filtering technique, an estimation method in control theory, to estimate drivers' VOTs in real time. Zhang et al. (2008) modeled the lane choice behavior of vehicles with a logit model. Then, they applied a piecewise feedback control method, which is based on different speeds on the HOT and GP lanes, to calculate the proportion of SOVs choosing HOT lanes. And the price was estimated backward from the logit model. Based on the self-learning method proposed by (Yin and Lou, 2009), Lou et al. (2011) considered the impacts of lane-changing behaviors with a multi-lane hybrid traffic flow model. Michalaka et al. (2011) formulated a robust pricing optimization problem to maximize the total throughput while maintaining the congestion on the HOT lanes. The traffic dynamics were described by the cell transmission model (Daganzo, 1994), and the flow-rates on the GP and HOT lanes

were estimated by a logit model.

However, there are some deficiencies in literature. Zhang et al. (2008) didn't consider the VOTs in the lane choice model. And VOTs are assumed to be known to the operators before determining the dynamic pricing schemes (Dorogush and Kurzhanskiy, 2015; Tan and Gao, 2018) assumed that. The feedback method in (Yin and Lou, 2009) can be regarded as an application of the ALINEA strategy (Papageorgiou et al., 1991), which is widely used in ramp metering. However, when the overall traffic is congested, a single Integral controller cannot keep zero queue on the HOT lanes. Examples can be found in Fig. 4, 5 and 8 in (Yin and Lou, 2009). In the self-learning method, a substantial residue queue exists on the HOT lanes in the simulation results for the self-learning method. At the same time, a theoretic study on the stability of the HOT lanes' operation is not available. In summary, existing pricing strategies cannot guarantee that the closed-loop system converges to the optimal state, in which the HOT lanes capacity is fully utilized but there is no queue on the HOT lanes, and a well-behaved estimation and control method is quite challenging and still elusive.

2.3 Lane Choice Models

If drivers choose the HOT lanes, they need to pay a toll, but at the same time they will save some travel time and improve the reliability of their trip. To develop more efficient pricing schemes for HOT lanes, it is critical to have a better understanding of the characteristics of such a traffic system, including how SOVs respond to the change in the price.

Generally, the utility for a driver is expressed by (2.1):

$$U_{in} = V_i(t_{in}, u_{in}, c_{in}, x_{in}) + \epsilon_{in}, \tag{2.1}$$

where t, v, and c are the measures of travel time, variability in travel time, and cost, respectively, for each route i, and traveler n. x is a vector of observable socioeconomic characteristics such as age, gender, annual household income, language spoken at home, wage rate, education, and other characteristics like flexibility of work arrival times and car occupancy; and ϵ_{in} is a random utility component (Lam and Small, 2001).

In the literature, several types of models have been proposed to capture the lane choice behaviors subject to different travel times, VOTs, and congestion prices. A nested logit model with the route choice, mode choice and transponder choice was used in (Sullivan, Edward, 2000). A binomial logit model was applied in (Lam and Small, 2001; Yin and Lou, 2009). When studying the potential SOV demand for traveling on HOT lanes using stated preference (SP) and revealed preference (RP) data from the travelers on the regular lanes at the Katy Freeway and Northwest Freeway corridors in Texas, Burris and Xu (2006) implemented a nested multinomial logit model to predict travelers mode choices. Gardner et al. (2013) presented three models to calculate the proportion of SOVs switching to the HOT lanes. The first model was an all-or-nothing assignment, in which all vehicles were assigned to the lane with lower generalized cost. The second model was a logit model, with the utility of each lane being the sum of travel time, congestion price and an independent and identically distributed Gumbel disturbance term. In the third model, the proportion of SOVs choosing the HOT lanes was exactly the proportion of travelers whose VOT exceeded the ratio of price and travel time difference. The last model can be regarded as an application of the user equilibrium principle, considering a Burr distribution for VOTs.

Although there are many models that can be used to capture travelers' lane choices, travel time and cost are always considered. These two factors indicate how much motorists value their time and how much they are willing to pay to switch lanes. In general, more SOVs want to pay and switch to the HOT lanes with a lower price when the overall system is congested. When the GP lanes are becoming more congested, there will be a larger travel

time difference between the GP and HOT lanes, and more SOVs will choose the HOT lanes; in this case, the operators need to increase the price to keep the HOT lanes from being congested [1]. So, for a corridor with HOT lanes, intuitively, SOVs' choices between the HOT and GP lanes depend on the difference in the travel times on the HOT and GP lanes, as well as congestion prices.

2.4   Value of Time

On the HOT lanes, the generalized cost of an SOV equals the product of the travel time and value of time (VOT) plus the toll; but on the GP lanes, it equals the product of the travel time and VOT. In economics, the VOT represents the opportunity cost of the time that a traveler spends on trips. The concept of VOT plays an important role in congestion pricing analysis as it shows user's trade-off between cost and time (Yang and Huang, 2005). For example, at the same price, SOVs with higher VOTs will be more willing to pay and switch to HOT lanes, and fewer SOVs will pay to use HOT lanes if the price is too high.

In 1997, the U.S. Department of Transportation (USDOT) developed and published its first manual for the valuation of travel time in economic analysis to be used by analysts in studies related to travel time and cost. The manual recommended different VOTs for different trip purposes, transportation modes, trip lengths, and vehicle operators. The latest manual was updated in 2016 (U.S. Department of Transportation, 2016). Details are shown in Table 2.1.

---

[1] A positive demand elasticity to the dynamic price for the HOT lanes is found in (Liu et al., 2011; Janson and Levinson, 2014). But, as more SOVs enter the HOT lanes and the tolls increase, there is a positive correlation between price and usage (Brent and Gross, 2018). In addition, from an economic view, consumers prefer to purchase the same amount of time savings at a lower price. For a more general case, we believe that consumers prefer to purchase the same amount of time savings at a lower price.

Table 2.1: Recommended Hourly VOT savings (2015 U.S. $per person-hour)

Category	Surface modes (except high-speed railway)		Air and high-speed rail travel	
	Low	High	Low	High
Local travel-				
Personal	9.50	16.30		
Business	20.30	30.50		
All purpose	10.00	17.00		
Intercity travel-				
Personal	16.30	24.50	31.00	46.50
Business	20.30	30.50	50.60	75.80
All purpose	17.20	25.80	38.90	58.30
Truck drivers	21.80	32.70		
Bus drivers	22.70	34.00		
Transit rail operators	36.90	55.30		
Locomotive engineers	33.30	49.90		
Airline pilots and engineers	69.40	104.10		

Traditionally, VOTs have been estimated offline. Small (1982) applied a discrete choice model with VOT included in the utility function. Then, the travelers VOT was estimated as the marginal rate of substitution between cost and time. The average VOT for work travel is $4.8/h$, which is assumed to be half of the gross wage rate. However, it can vary from 20 to 100 percent depending on the city and population group. With the maximum log-likelihood criterion, Lam and Small (2001) estimated the mean VOT based on the survey data and

19

loop detector data on SR-91. The VOT is estimated by (2.2):

$$VOT_n = \frac{\partial V}{\partial t_n} / \frac{\partial V}{\partial c_n}, \tag{2.2}$$

where n indicates the travel, V represents the utility, t_n and c_n are the traveler's travel time and cost respectively. Later, Brownstone et al. (2003) estimated the median VOT of $30/h$ with the revealed preference data from drivers, loop detector data and ETC data from I-15 in Southern California. The upper quartile of the distribution is to $43/h$ and the lower is $23/h$.

However, those estimation methods cannot be applied in real-time operation, since it takes time to collect data from FasTrak users. At the same time, Liu et al. (2007) found that the median VOT is dependent of the departure time during the peak hours with the data from SR-91. Tseng and Verhoef (2008) showed that VOTs vary by time of day with the stated-preference survey data from Dutch commuters . Campbell (2016) mentioned that it is difficult to accurately quantify the true benefits of dynamic pricing projects without a quick estimation of VOTs. Based on those observations, we conclude that it is necessary to estimate VOTs in real-time to determine effective pricing strategies.

2.5   Departure Time Choices

In the real world, HOT and GP lanes can both become congested during the peak hours due to the high travel demand. On May 15th, 2019, the speed was about 32 mph on the innermost SR-91 Express Lane, and 33 mph on the second lane; the corresponding occupancies were both above 11% at 6pm; the average speed on the GP lanes was 20 mph, and the occupancy was about 30% (California Department of Transportation, 2019). This phenomenon indicates that, even with pricing strategies, the HOT lanes cannot be operated at the optimal state

during peak periods.

It has long been known that drivers' departure time choice behavior is one fundamental cause of congestion. Departure time choice and route choice behaviors determine the spatial and temporal distribution of morning/evening commuting trips. Vickrey (1969) first introduced the departure time choice problem for a single bottleneck. When the system reaches the equilibrium state, "the journey costs at all departure times actually used are equal, and less than those which would be experienced by a single vehicle at any unused time" (Wardrop, 1952). So, when all drivers are identical, the costs are constant over the congested period (Arnott et al., 1987). Some follow-up studies for a single bottleneck can be found in (Hendrickson and Kocur, 1981; Arnott et al., 1990a).

Later, the departure time choice problem has been studied in more complex transportation networks. For example, the situation where one O/D pair is connected by two parallel routes was studied in (Mahmassani and Herman, 1984; Arnott et al., 1990b). The departure time choice problem for a two-tandem bottleneck scenario was investigated in (Kuwahara, 1990; Yang and Meng, 1998). The simultaneous departure time and route choice model for multiple O/D pairs was formulated through a variational inequality problem in (Wie et al., 1995; Szeto and Lo, 2004). In the presence of pricing strategies, Vickery's model was used to analyze the departure time choice for parallel free and toll roads in (Arnott et al., 1990b; De Palma and Lindsey, 2000). A mathematical program with equilibrium constraints formulation was applied to solve the dynamic route-departure time choice and determine the step tolls in (Joksimovic et al., 2005; Viti et al., 2003). However, to our knowledge, the departure time choice for HOT lanes was only studied by (Boyles et al., 2015). In this paper, they assumed the demand of HOVs was lower than the capacity of the HOT lane, then some "strategic" SOVs can choose their departure time and lane group in the presence of pricing schemes. They derived the equilibrium departure time profile with the variational inequality, and found an approximate solution with the method of successive average. In the numerical

study, they chose four metrics to measure the performance of four pricing schemes. Results showed that all toll algorithms resulted in lower average travel time than the no-toll setup.

Chapter 3

A Control Theoretic Approach to Simultaneously Estimate Average Value of Time and Determine Dynamic Price for High-occupancy Toll Lanes

3.1 Introduction

This study is inspired by (Yin and Lou, 2009), in which a simultaneous estimation and control problem is studied for a freeway with the HOT and GP lanes. Note that the simultaneous estimation and control problems are relatively new in the transportation field but have been studied in many other areas. In the economics, Taylor (1970) applied a Kalman filter to estimate the firm's long-term demand for inventoried goods when a distributed lags

exists. After the demand was estimated, a feedback decision rule was applied to minimize the expected cost. An estimation and identification of parameters and control variables of electric-motor-driven motion system can be found in (Ohnishi et al., 1994). In the field of space science, Habib (2013) adopted an extended Kalman filter (EKF) to estimate the position and velocity of a spacecraft, and then designed a proportional - derivative (PD) controller to control the system.

In general, the HOT systems are stochastic, because of the traffic demands, the distribution of SOVs' VOTs and the traffic flow models. The stochastic traffic demands were considered in (Yin and Lou, 2009; Lou et al., 2011). A Burr-distributed VOTs was introduced to incorporate heterogeneous drivers in (Gardner et al., 2013). In this study, the limited impact of stochastic demands is numerically proved in Section 3.7. This result help justify the importance of the stability property of our method.

Table 3.1 is the list of notations in this study.

Table 3.1: List of notations

Variables	Definitions	Variables	Definitions
$q_1(t)$	Demands of HOVs	$q_2(t)$	Demands of SOVs
$q_3(t)$	Demands of paying SOVs	C_1	Capacity of the HOT lanes
C_2	Capacity of the GP lanes	$\zeta(t)$	Residual capacity of the HOT lanes
$\lambda_1(t)$	Queue length on the HOT lanes	$\lambda_2(t)$	Queue lengths on the GP lanes
ϵ	Infinitesimal positive number	Δt	Time step size
$w_1(t)$	Queuing time on the HOT lanes	$w_2(t)$	Queuing time on the GP lanes
$w(t)$	Queuing time difference between the GP and HOT lanes	$g_1(t)$	Throughput of the HOT lanes
$g_2(t)$	Throughput of the GP lanes	$u(t)$	Time-dependent price for paying SOVs
$\pi^\star$	True average VOT	$\pi(t)$	Estimated average VOT
K_1	Integral controller coefficient for $\lambda_1(t)$	K_2	Integral controller coefficient for $\zeta(t)$

In this study, we want to: (1) define a new variable called the residual capacity and present a simpler formulation of the point queue model; (2) propose a simple feedback control theoretic approach to estimate the average value of time and calculate the dynamic price; and (3) analytically and numerically prove that the closed-loop system is stable and guaranteed to converge to the optimal state. The rest of this study is organized as follows. In Section 3.2, we define a new variable: the residual capacity of the HOT lanes. Then, we describe the system dynamics for a freeway with the HOT and GP lanes based on the residual capacity. We also define the simultaneous estimation and control problem for the HOT lanes. In Section 3.3, we provide the solution of the control problem with constant demand, assuming the operators have full information of the true average VOT of SOVs before designing the pricing schemes. In Section 3.4, we provide a detailed review of two control methods in (Yin and Lou, 2009). In Section 3.5, we present a control theoretic approach to estimate the average VOT of SOVs, and calculate the dynamic price with the estimated average VOT and the travel time difference. In Section 3.6, we analyze the equilibrium state and stability of the closed-loop control system. In Section 3.7, we show that our method is effective with numerical examples, and compare its performance with the methods in (Yin and Lou, 2009). Then, we check the robustness of the control system with respect to variations in the demands of SOVs and HOVs. At the same time, we numerically prove that the closed-loop system is stable and guaranteed to converge to the optimal state. We also examine the effect of the scale parameter in the lane choice model. In Section 3.8, we conclude the study and provide future research topics.

3.2 Definitions, system dynamics, and problem statement

The study site is a freeway corridor with an origin and a destination, as shown in Figure 3.1. The downstream of both lanes is uncongested initially. The freeway has two types of lanes: the HOT and GP lanes. There is one bottleneck on the GP lanes because of lane drop, but not on the HOT lanes. The operation principle for the HOT lanes are as follows:

1. HOVs can use the HOT lanes for free.

2. SOVs have to pay a price to use the HOT lanes.

Thus, SOVs can be put in two categories: (1) paying SOVs that pay a price to use the HOT lanes; and (2) non-paying SOVs that stay on the GP lanes all the time. Therefore, the HOVs and paying SOVs share the HOT lanes, but the non-paying SOVs use the GP lanes. Two sets of detectors are installed on the freeway to detect the demands on the HOV, HOT and GP lanes.
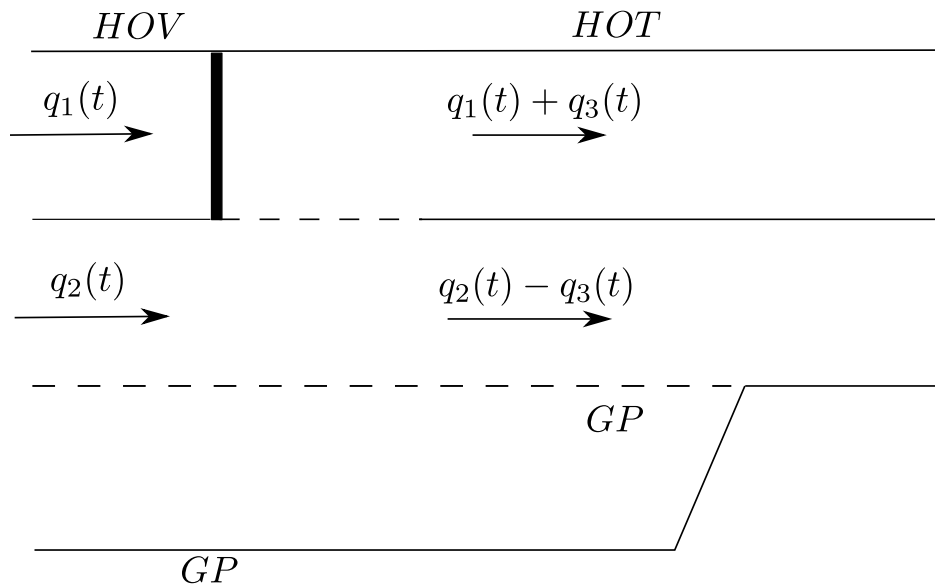
Figure 3.1: Illustration of the traffic system with HOT lanes.

26

3.2.1 Definitions of variables

We describe the road and traffic characteristics with the following variables:

- Traffic demands: $q_1(t)$ and $q_2(t)$ are demands of HOVs and SOVs, respectively.

- Capacities: C_1 and C_2 are the capacities of the HOT and GP lanes, respectively.

- Queue lengths: $\lambda_1(t)$ and $\lambda_2(t)$ are the queue lengths on the HOT and GP lanes, respectively.

- Queuing times: $w_1(t)$ and $w_2(t)$ are the queuing times for vehicles leaving at t on the HOT and GP lanes, respectively; $w(t) = w_2(t) - w_1(t)$ is the queuing time difference.

In this study, we have the following two assumptions regarding the demand patterns. First, the total demand exceeds the total capacity during the study period $t \in [0, T]$; i.e.,

$$\int_0^T (q_1(s) + q_2(s))ds > (C_1 + C_2)T;$$

therefore, the overall traffic system is congested. Second, the demand of HOVs is below the capacity of the HOT lanes, i.e.,

$$q_1(t) < C_1. \tag{3.1}$$

Those assumptions are consistent with the findings in (Kwon and Varaiya, 2008). By introducing the HOT lanes, however, we can charge some SOVs to use the underutilized HOV lanes. We denote the demand of paying SOVs by $q_3(t)$, and the demands on the GP and HOT lanes should be $q_2(t) - q_3(t)$ and $q_1(t) + q_3(t)$, respectively. Then, we define the residual capacity of the HOT lanes by

$$\zeta(t) = C_1 - q_1(t) - q_3(t). \tag{3.2}$$

Meanwhile. we denote the time-dependent price for paying SOVs by $u(t)$.

3.2.2   Models of system dynamics

Traffic flow model

We apply the point queue model (PQM) to model traffic dynamics (Vickrey, 1969; Jin, 2015). In the point queue model, all drivers follow the first-in-first-out (FIFO) rule, and the travel time is the sum of the free-flow travel time and the queuing time. For vehicles in a homogeneous freeway segment, their travel time difference only exists in the queuing time on the HOT and GP lanes.
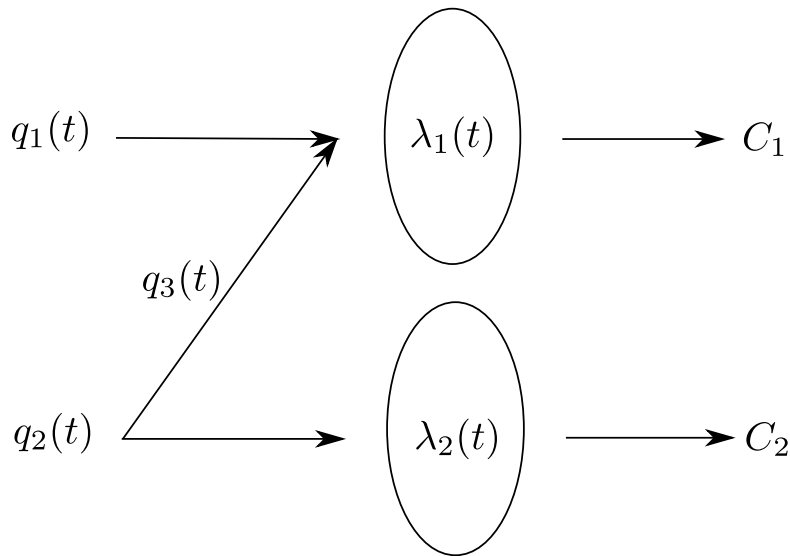
Figure 3.2: Point queue representation of the traffic system with HOT lanes.

For the HOT and GP lanes in this system (see Figure 3.2), the dynamics of the two point

queues are described by the following ordinary differential equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\lambda_1(t) = \max\{q_1(t) + q_3(t) - C_1, -\frac{\lambda_1(t)}{\epsilon}\}, \tag{3.3a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\lambda_2(t) = \max\{q_2(t) - q_3(t) - C_2, -\frac{\lambda_2(t)}{\epsilon}\}. \tag{3.3b}$$

With the definition of $\zeta(t)$ in (3.2), we rewrite (3.3) as

$$\frac{\mathrm{d}}{\mathrm{d}t}\lambda_1(t) = \max\{-\zeta(t), -\frac{\lambda_1(t)}{\epsilon}\}, \tag{3.4a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\lambda_2(t) = \max\{q_1(t) + q_2(t) - C_2 - C_1 + \zeta(t), -\frac{\lambda_2(t)}{\epsilon}\}, \tag{3.4b}$$

where $\epsilon = \lim_{\Delta t \to 0^+} \Delta t$ is an infinitesimal number and equals Δt in the discrete form. Then, the discrete version of the PQM becomes

$$\lambda_1(t + \Delta t) = \max\{-\zeta(t)\Delta t + \lambda_1(t), 0\}, \tag{3.5a}$$

$$\lambda_2(t + \Delta t) = \max\{(q_1(t) + q_2(t) - C_2 - C_1 + \zeta(t))\Delta t + \lambda_2(t), 0\}. \tag{3.5b}$$

From (3.4a), we can see that the queue changing rate on the HOT lanes is determined by either the existing queue length or the residual capacity.

In addition, the throughputs of the HOT and GP lanes are given by

$$g_1(t) = \min\{C_1 - \zeta(t) + \frac{\lambda_1(t)}{\epsilon}, C_1\}, \tag{3.6a}$$

$$g_2(t) = \min\{q_1(t) + q_2(t) - C_1 + \zeta(t) + \frac{\lambda_2(t)}{\epsilon}, C_2\}. \tag{3.6b}$$

So, the queuing times on the HOT and GP lanes are

$$w_1(t) = \frac{\lambda_1(t)}{C_1}, \tag{3.7a}$$

$$w_2(t) = \frac{\lambda_2(t)}{C_2}. \tag{3.7b}$$

Furthermore, the queuing time difference, denoted by $w(t)$, is

$$w(t) = w_2(t) - w_1(t) = \frac{\lambda_2(t)}{C_2} - \frac{\lambda_1(t)}{C_1}. \tag{3.8}$$

Lane choice model

In this study, we assume a logit model to calculate the proportion of SOVs choosing the HOT lanes, $Pr(t)$:

$$Pr(t) = \frac{\exp(\alpha^* V_{HOT}(t))}{\exp(\alpha^* V_{HOT}(t)) + \exp(\alpha^* V_{GP}(t))},$$

where $V_{HOT}(t)$ and $V_{GP}(t)$ represent the measurable utility of drivers using the HOT and GP lanes at time step t; and α^* is a scale parameter, which determines the variation of a Gumbel distribution (Train, 2009). We assume that the drivers have full information of the travel time and dynamic price. We denote the free-flow travel time by t_f. Then, the measurable utility for paying SOVs is $-(u(t) + \pi^*(t_f + w_1(t)))$, and that for non-paying SOVs is $-\pi^*(t_f + w_2(t))$, where π^* represents the true VOT. It is obvious that the VOT is not affected by the scale parameter. Then, the lane choice model for SOVs is simplified as

$$Pr(t) = \frac{1}{1 + \exp(\alpha^*(u(t) - \pi^* w(t)))}. \tag{3.9}$$

The total demand of SOVs is $q_2(t)$, and $w(t)$ is calculated by (3.8). Then, the demand of

paying SOVs is

$$q_3(t) = \frac{q_2(t)}{1 + \exp(\alpha^*(u(t) - \pi^*w(t)))}. \qquad (3.10)$$

With the definition of $\zeta(t)$, (3.10) is rewritten as

$$\zeta(t) = C_1 - q_1(t) - \frac{q_2(t)}{1 + \exp(\alpha^*(u(t) - \pi^*w(t)))}. \qquad (3.11)$$

3.2.3   Simultaneous estimation and control problem

In this study, we only consider the first two objectives for operating the HOT lanes, which aim to optimize the whole system's performance without sacrificing the HOT lanes' LOS (Perez and Sciara, 2003): (1) keeping zero queue on the HOT lanes, and (2) maximizing the HOT lanes' throughput. Since the GP lanes are always congested and their throughput is fixed at capacity, maximizing the HOT lanes' throughput is the same as maximizing the whole system's throughput in this case.

When both of the above objectives are met at the same time, we refer to the traffic state as an optimal state. That is, in an optimal state, we have: (1)

$$\lambda_1(t) = 0, \qquad (3.12)$$

and (2)

$$g_1(t) = C_1. \qquad (3.13)$$

Lemma 3.2.1. *In the optimal state, the residual capacity on the HOT lanes is given by*

$$\zeta(t) = 0. \tag{3.14}$$

Proof: Since $\lambda_1(t) = 0$ in the optimal state, (3.6a) is equivalent to

$$g_1(t) = \min\{C_1 - \zeta(t), C_1\}.$$

The maximum throughput of the HOT lanes occurs if and only if $\zeta(t) \leq 0$. Also, $\zeta(t)$ cannot be negative in the optimal state from (3.4a). So, at the optimal state, $\zeta(t) = 0$. $\quad\square$

Considering the different choice behavior between different drivers, the operators cannot simply force SOVs to switch into or out of the HOT lanes to achieve the operation objectives. As in (Yin and Lou, 2009), we are interested in finding an appropriate pricing scheme that can drive the system to the optimal state, since the price influence SOVs' lane choice behavior according to (3.10), which in turn would influence the HOT lanes' queue size according to (3.3). In this sense, the congestion price serves as an actuation signal to the system. In addition, we are interested in estimating the average VOT of SOVs simultaneously since the VOT is a key parameter in any lane choice models for the HOT lanes studies.

3.3   Solution of the control problem with constant demand

We consider a simple case when the operators know the true average VOT of SOVs, then the problem is simplified as a control problem. We assume the demand of HOVs and SOVs are time-independent, that is, the demand of HOVs is constant at $q_1 < C_1$, and the demand of SOVs is constant at $q_2 > C_2 + C_1 - q_1$. These are consistent with the two assumptions

regarding the demand patterns in Section 3.2. In this demand pattern, if we do not allow SOVs to use the HOT lanes, then the GP lanes are congested, while the HOT lanes are underutilized.

With appropriate pricing schemes, the system reaches the optimal state when the two operational objectives stated in Section 3.2.3 are met at the same time. That is, (1)

$$\lambda_1(t) = 0;$$

i.e. there is no queue on the HOT lanes; and (2)

$$\zeta(t) = 0;$$

i.e., there is no residual capacity on the HOT lanes.

Initially both lanes are uncongested, i.e., $\lambda_1(0) = \lambda_2(0) = 0$. The demand on the HOT lanes equals the capacity, and the demand on the GP lanes is $q_2 - (C_1 - q_1)$. According to (3.4a), the queue changing rate is zero, so there will be no queue on the HOT lanes, and the queuing time on the HOT lanes is 0. Based on (3.5) and (3.7), the queue length on the GP lanes is $\lambda_2(t) = (q_2 + q_1 - C_1 - C_2)t$, and the queuing time on the GP lanes is $w_2(t) = \frac{q_1 + q_2 - C_1 - C_2}{C_2}t$. Then $w(t) = w_2(t) = \frac{q_1 + q_2 - C_1 - C_2}{C_2}t$. According to (3.11), the lane choice model is written as

$$C_1 - q_1 - \frac{q_2}{1 + \exp(\alpha^*(u(t) - \pi^* \frac{q_1 + q_2 - C_1 - C_2}{C_2}t))} = 0,$$

which leads to

$$u(t) = \frac{q_1 + q_2 - C_1 - C_2}{C_2}\pi^*t + \frac{\ln \frac{q_1 + q_2 - C_1}{C_1 - q_1}}{\alpha^*}. \tag{3.15}$$

This analytical result reveals that the price should increase linearly with constant demand

33

pattern and known true average VOT. In the following sections, we assume that the operators know the SOVs follow a logit model with a scale parameter of α^* for the lane choice, but they do not know the true average VOT. In this sense, we need to estimate the average VOT to design appropriate pricing schemes for the HOT lanes. We denote $\pi(t)$ as the estimated average VOT. Then, the operators should replace π^* by $\pi(t)$ in (3.15) to calculate the price.

3.4 Review of Yin and Lou's methods

As mentioned in the beginning of this chapter, this study is inspired by (Yin and Lou, 2009). We think it is necessary to provide a review of this paper before introducing our method. In (Yin and Lou, 2009), they proposed two dynamic pricing strategies for operating HOT lanes to provide an uncongested traffic condition on the HOT lanes while maximizing the freeway's throughput: the feedback method, and the self-learning method.

3.4.1 Feedback method

In the feedback method, one loop detector was required downstream of the toll reader to detect the occupancy on the HOT lanes. The dynamic price on HOT lanes, $u(t)$, was calculated by an Integral controller:

$$u(t + \Delta t) = u(t) + K_I(O_{HOT}(t) - O^*_{HOT}(t)),$$

where $O_{HOT}(t)$ was the measured occupancy on the HOT lanes at time step t, $O^*_{HOT}(t)$ was the desired occupancy, and was usually equal or slightly less than the critical occupancy. The error term was the difference between the desired occupancy and the measured occupancy. And K_I was the coefficient for the Integral controller. The framework for the feedback method is illustrated in Figure 3.3a. Since the occupancy is not a state variable in the PQM,

so they replace the occupancy with the demand of the HOT lanes in the simulation. The error term was then the difference between the actual and desired demand on the HOT lanes. Then, the control logic becomes

$$u(t + \Delta t) = u(t) + K_I(q_{HOT}(t) - q^*_{HOT}(t)).$$ (3.16)

Based on this control law, when $q_{HOT}(t) > q^*_{HOT}(t)$, the price would increase; when $q_{HOT}(t) < q^*_{HOT}(t)$, the price would decrease; and when $q_{HOT}(t) = q^*_{HOT}(t)$, the price would be constant.

Even though the feedback method is simple to implement, it fails to achieve the two objectives by directly calculating the price. Let's consider the same demand pattern in Section 3. When the two objectives are met at the same time, $q_{HOT}(t) = q^*_{HOT}(t)$, then the price would be constant according to (3.16). According to (3.10), more SOVs are willing to pay and switch, thus the HOT lanes would be congested. So, the control system is unstable.

3.4.2 Self-learning method

In the self-learning method, two sets of loop detectors were required. The first set of detectors was installed before the toll-tag reader to detect the demand on the HOV and GP lanes; while the second set of detectors was installed after the reader to detect the demand on the HOT and GP lanes. A logit model was applied to describe SOV's lane choice and a PQM was used to capture traffic dynamics. In each step, they first learned SOV drivers' willingness to pay (WTP) by mining the loop detector data, and then determined the price based on the demand, estimated travel times, and the calibrated WTP. When calibrating the WTP, a discrete Kalman filter was incorporated to estimate the three parameters in the logit model. The framework for the self-learning method is shown in Figure 3.3b.

(a) The feedback method   (b) The self-learning method
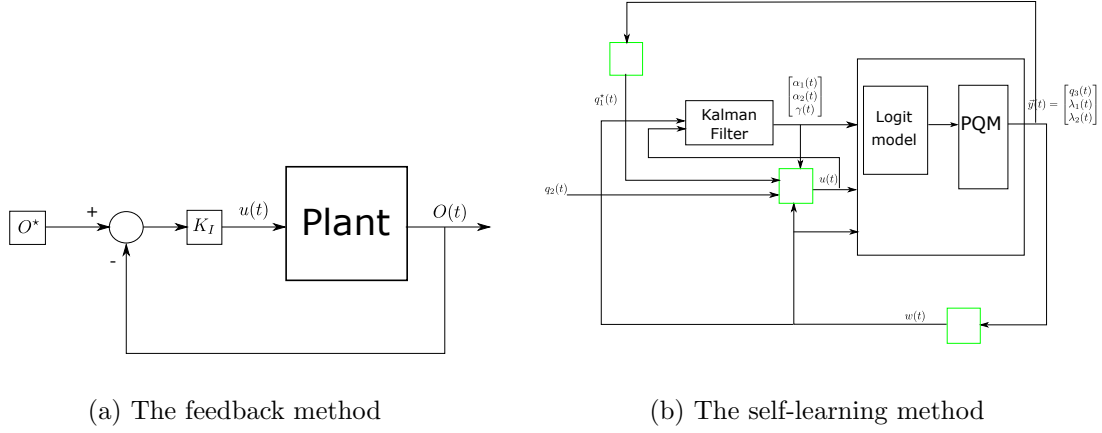
Figure 3.3: Block diagrams of two methods in (Yin and Lou, 2009).

In the logit model, α_1 and α_2 represented the marginal effect of travel time and prices on drivers' utility respectively, and γ captured other factors that affect the WTP. Clearly, α_1/α_2 was the trade-off between time savings and prices, i.e., VOT. In each step, they updated the Kalman gain and error covariance matrix in the Kalman filter. Then, they determined the dynamic price for the HOT lanes backward based on the logit model:

$$u(t) = \frac{\ln \frac{q_2(t) - q_{HOT}^*(t)}{q_{HOT}^*(t)} + \alpha_1(t)w(t) - \gamma(t)}{\alpha_2(t)}, \tag{3.17}$$

where $q_{HOT}^*(t)$ was the optimal demand on the HOT lanes, and $w(t)$ was the travel time difference on the GP and HOT lanes at time t.

When we consider the same constant demand as Section 3. In the optimal state, $q_{HOT} = C_1$. Then, the queue length on the GP lanes is $\lambda_2(t) = (q_2 + q_1 - C_1 - C_2)t$, and the travel time difference is $w(t) = \frac{q_1 + q_2 - C_1 - C_2}{C_2}t$. Based on (3.17), the price should be

$$u(t) = \frac{\ln \frac{q_2 - C_1}{C_1} + \alpha_1 \frac{q_1 + q_2 - C_1 - C_2}{C_2}t - \gamma}{\alpha_2}. \tag{3.18}$$

36

Since the price is calculated reversely from the logit model, and it only influences the lane choice for SOVs, $C_1 - q_1$ should replace C_1 in the log term in (3.18).

Note that there is an inconsistency in their simulation setup. In both scenario 1 and 2, the average demand of HOVs is 300 vph for the feedback method, but 600 vph for the self-learning method. This can be seen by comparing Figure 4a and 7a for scenario 1 and Figure 5a and 9a for scenario 2. Such an inconsistency can also be confirmed by comparing the corresponding queue lengths in the numerical results.

3.5   A new control theoretic approach

In this study, we are interested in simultaneously estimating the average VOT of SOVs and calculating the dynamic price for the HOT lanes when the operators want to achieve the operation objectives: (3.12) and (3.14). We want to combine the advantages of the two methods in (Yin and Lou, 2009). Similar to the self-learning method, we first estimate the average VOT, and then calculate the dynamic price. However, we will take advantage of the simplicity of the feedback controller, and incorporate it in the estimation process.

The block diagram of the control system is shown in Figure 3.4a. This is a feedback system, in which the price and the system dynamics are connected together such that each system influences the other and their dynamics are strongly coupled (Aström and Murray, 2010). Let's recall the objectives for operating the HOT lanes: $\lambda_1(t) = 0$ and $\zeta(t) = 0$. They are the reference signals of the system, $\vec{r}(t)$. Then, $\lambda_1(t)$ and $\zeta(t)$ are the two error signals related to the traffic condition on the HOT lanes. With appropriate pricing schemes, $u(t)$, we can achieve those objectives. The plant has a logit model and a point queue model, and the detailed block diagram is shown in Figure 3.4b. Inside the plant, the inputs for the lane choice model are $u(t)$, $q_2(t)$ and $w(t)$, and the output is $\zeta(t)$; for the traffic model, the inputs

are $q_1(t)$, $q_2(t)$ and $\zeta(t)$, and the outputs are $\lambda_1(t)$, $\lambda_2(t)$ and $w(t)$.

(a) Block diagram of the control system   (b) Block diagram of the plant
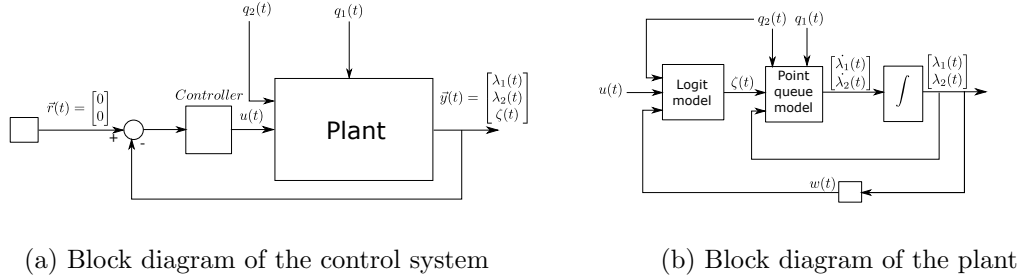
Figure 3.4: Block diagram of the control system.

3.5.1 A feedback estimation method

PID control is the most common way of using feedback in engineering systems so far. In traffic operation, PID control is widely used in ramp metering (Papageorgiou et al., 1991) and variable speed limits (Zhang et al., 2006). In this study, we want to apply it to solve the estimation problem. Since π^* is a constant value, we apply an Integral controller (I-controller) to estimate it. The operators can gradually estimate the average VOT of SOVs through learning their reaction to the dynamic price and traffic condition.

We assume that the operators have full information of the flow-rate and queue length on the HOT lanes. For time-dependent demand, the I-controller is designed as

$$\frac{\mathrm{d}}{\mathrm{d}t}\pi(t) = K_1\lambda_1(t) - K_2\zeta(t). \tag{3.19}$$

If there is a queue on the HOT lanes, i.e., $\lambda_1(t) > 0$, we should increase the estimated average VOT of SOVs. If the HOT lanes are underutilized, i.e., $\zeta(t) > 0$, we should decrease the estimated average VOT. Therefore, all the coefficients in (3.19), including K_1 and K_2, should be positive. In the optimal state, $\pi(t)$ would be constant since the error signals are zero.

3.5.2 Calculation of the dynamic price

With the estimated parameters in the previous subsection, we calculate the dynamic price for the HOT lanes as in (3.15), except that the true parameters and variables are replaced by the estimated values:

$$u(t) = \pi(t)w(t) + \frac{\ln \frac{q_1+q_2-C_1}{C_1-q_1}}{\alpha^*}. \tag{3.20}$$

Since $\pi(t) \neq \pi^*$ initially, we would expect some fluctuation in the traffic condition and the dynamic price even if the demand is constant.

Figure 3.5 illustrates the controller formed by (3.19) and (3.20). The controller is implemented in two steps: in the first it estimates the average VOT based on $\lambda_1(t)$ and $\zeta(t)$, and in the second it calculates the dynamic price.
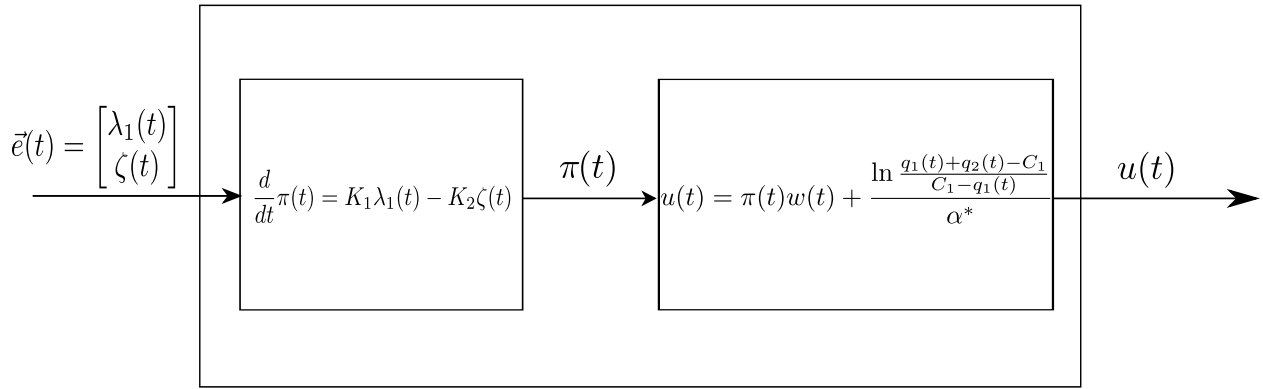
Figure 3.5: Block diagram of the controller.

3.6 Analytical properties of the closed-loop control system with constant demand

In this section, we study the properties of the closed-loop system with constant demand levels. The whole system system is formulated with the definition of $\zeta(t)$ in (3.2), the feedback estimator in (3.19) and the dynamic price in (3.20), the point queue models in (3.4) and the lane choice model in (3.11). There are six unknown variables in the system: $\lambda_1(t)$, $\lambda_2(t)$, $\zeta(t)$, $w(t)$, $\pi(t)$ and $u(t)$. Different from Section 3.3, here we assume we don't know the true average VOT of SOVs.

3.6.1 Equilibrium state

Further by substituting (3.20) into (3.11), we obtain the following system:

$$\frac{\mathrm{d}}{\mathrm{d}t}\lambda_1(t) = \max\{-\zeta(t), -\frac{\lambda_1(t)}{\epsilon}\}, \tag{3.21a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\lambda_2(t) = \max\{q_1 + q_2 - C_1 - C_2 + \zeta(t), -\frac{\lambda_2(t)}{\epsilon}\}, \tag{3.21b}$$

$$\zeta(t) = C_1 - q_1 - \frac{q_2}{1 + \exp(\alpha^*(u(t) - \pi^* w(t)))}, \tag{3.21c}$$

$$w(t) = \frac{\lambda_2(t)}{C_2} - \frac{\lambda_1(t)}{C_1}, \tag{3.21d}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\pi(t) = K_1\lambda_1(t) - K_2\zeta(t), \tag{3.21e}$$

$$u(t) = \pi(t)w(t) + \frac{\ln \frac{q_1 + q_2 - C_1}{C_1 - q_1}}{\alpha^*}. \tag{3.21f}$$

We define the equilibrium state as when $\dot{\lambda}_1(t) = 0$ and $\dot{\pi}(t) = 0$, which leads to $\lambda_1(t) = \lambda_1$ and $\pi(t) = \pi$. From (3.21a) and (3.21e) we can see that, $\lambda_1 = 0$, and $\zeta(t) = \zeta = 0$. From (3.21b) we find that, in the equilibrium state, $\frac{\mathrm{d}}{\mathrm{d}t}\lambda_2(t) = q_1 + q_2 - C_1 - C_2$. At a very large time t, the equilibrium queue length on the general purpose lanes is $\lambda_2(t) \approx (q_1 + q_2 - C_1 - C_2)t$.

3.6.2 Stability of the equilibrium state

In this subsection, we analyze the local stability of the equilibrium state at a large time t, when $\lambda_1(t)$ and $\zeta(t)$ are both very small. By substituting (3.21c) into (3.21e), we obtain

$$- \frac{\dot{w}(t)}{\alpha^* w^2(t)} \ln(\frac{q_1 + q_2 - C_1 + \zeta(t)}{C_1 - q_1 - \zeta(t)} \frac{C_1 - q_1}{q_1 + q_2 - C_1}) + \frac{1}{\alpha^* w(t)} \frac{q_2}{(q_1 + q_2 - C_1 + \zeta(t))(C_1 - q_1 - \zeta(t))} \frac{\mathrm{d}}{\mathrm{d}t} \zeta(t)$$
$$= K_1 \lambda_1(t) - K_2 \zeta(t). \tag{3.22}$$

Since $\lambda_2(t)$ is very large, and $\zeta(t)$ is very small near the equilibrium state, from the above equation we have the following approximate dynamics for $\zeta(t)$:

$$\frac{\mathrm{d}}{\mathrm{d}t} \zeta(t) \approx \beta t (K_1 \lambda_1(t) - K_2 \zeta(t)), \tag{3.23}$$

where $\beta = \frac{\alpha^*(q_1 + q_2 - C_1 - C_2)(q_1 + q_2 - C_1)(C_1 - q_1)}{C_2 q_2} > 0$. Therefore, the system of (3.21a) and (3.23) approximates the dynamics of the original closed-loop control system, (3.21), at the equilibrium state subject to a small disturbance after a long time. Clearly the equilibrium state of the approximate system is at $(\lambda_1^*, \zeta^*) = (0, 0)$. Note that (3.23) is a liner time-variant system; thus the approximate system is a switching linear time-variant system.

Theorem 3.6.1. *The approximate system of (3.21a) and (3.23) is locally stable at the equilibrium state $(0, 0)$ after a long time. Thus the closed-loop system, (3.21), is locally stable at its equilibrium state.*

Proof: From (3.21a), there are two phases for the queue dynamics on the HOT lanes.

1. When $\zeta(t) \geq \frac{\lambda_1(t)}{\epsilon}$, $\lambda_1(t + \epsilon) = 0$, and the queue on the HOT lanes vanishes. In this case, (3.23) can be simplified as

$$\frac{\mathrm{d}}{\mathrm{d}t} \zeta(t) \approx -\beta K_2 t \zeta(t).$$

The solution of the above equation is

$$\zeta(t) \approx \zeta(0)e^{-\frac{1}{2}\beta K_2 t^2}, \tag{3.24}$$

which converges to the equilibrium value 0 in a Gaussian manner, much faster than exponentially.

2. When $\zeta(t) < \frac{\lambda_1(t)}{\epsilon}$, $\lambda_1(t) > 0$, and the queue length on the HOT lanes is positive. In this case, (3.21a) can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\lambda_1(t) = -\zeta(t).$$

At a very large time t, $\frac{\mathrm{d}}{\mathrm{d}t}\zeta(t)$ is finite, and (3.23) is equivalent to

$$K_1\lambda_1(t) - K_2\zeta(t) \approx \frac{1}{\beta t}\frac{\mathrm{d}}{\mathrm{d}t}\zeta(t) \to 0.$$

Thus

$$\zeta(t) \approx \frac{K_1}{K_2}\lambda_1(t) \approx \frac{K_1}{K_2}\lambda_1(0)e^{-\frac{K_1}{K_2}t}. \tag{3.25}$$

which converges to the equilibrium state exponentially.

In both cases, the approximate system is stable. □

We conclude that there are two convergence patterns of the approximate model after a long time:

Pattern 1: When $\lambda_1(t) = 0$, $\zeta(t)$ converges in a Gaussian manner.

Pattern 2: Both $\lambda_1(t)$ and $\zeta(t)$ converge exponentially, and $\frac{\lambda_1(t)}{\zeta(t)} = \frac{K_2}{K_1}$.

3.7 Numerical results

In this section, we provide numerical results of the three methods discussed above. The site is a freeway segment with lane drop downstream of the GP lanes (see Figure 3.1), and the capacity for one HOT and one GP lane is 30 veh/min. The downstream is not congested initially ($\lambda_1(0) = \lambda_2(0) = 0$). The study period is 20 minutes, and the time-step size is $1/60$ min. For simplicity, We assume the true average VOT is \$0.5/min ($\pi^* = $ \$0.5/min), and $\alpha^* = 1$. Our initial guess of average VOT is \$0.25/min.

3.7.1 Comparison of three controllers with constant demand

We first consider a constant demand profile. The demand of HOVs is constant at $q_1(t) = 10$ veh/min, and the demand of SOVs is constant at $q_2(t) = 60$ veh/min. Then $q_3^*(t) = 20$ veh/min.

For the controller, (3.19), we set $K_1 = $ \$0.1 /min^2 and $K_2 = $ \$0.1/min.
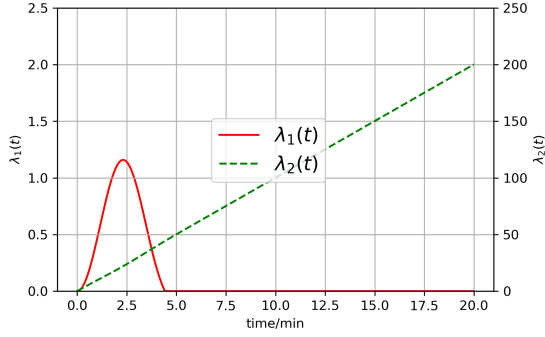
(a) Queue lengths
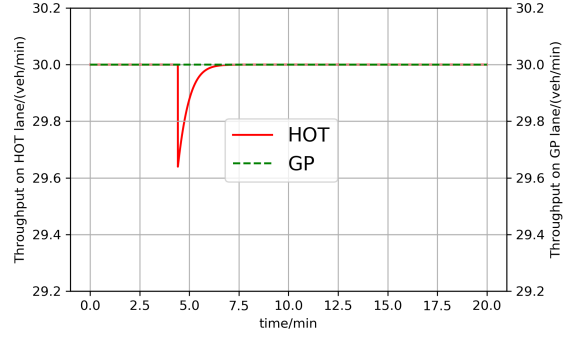
(b) Throughputs
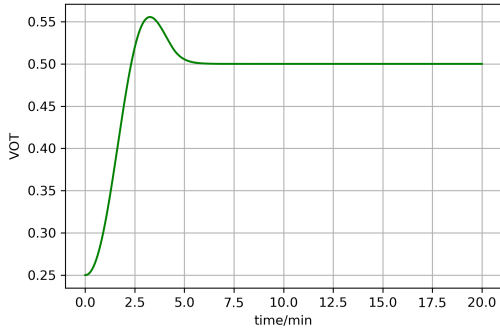
(c) Estimated average VOT
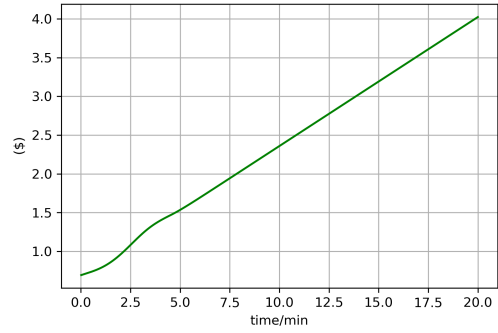
(d) Price

Figure 3.6: Numerical results of our method ((3.19) and (3.20)) with $K_1 = \$0.1/\text{min}^2$ and $K_2 = \$0.1/\text{min}$

.

In Figure 3.6a, the queue on the HOT lanes becomes zero quickly, despite some fluctuations at the beginning. The queue on the GP lanes increases linearly with time after the queue is eliminated on the HOT lanes. The HOT lanes are underutilized initially, but later the throughput is at capacity, as shown in Figure 3.6b. The average throughput of the HOT lanes is 29.96 veh/min, and the throughput of the GP lanes is 30 veh/min since the GP lanes are always congested. Since there is no residual queue and no residual capacity on the HOT lanes, the control system can reach the optimal state after some time. In Figure 3.6c, after about 6 minutes, the true the true average VOT ($0.5/min) is estimated. In Figure 3.6d, the dynamic price increases linearly with time after the true average VOT is estimated, and

the price is \$4.024 at 20 minutes.

In order to check the effectiveness of our method, we go back to the analytical results. From (3.15), we get $u(t) = \frac{1}{6}t + \ln 2$. At $t = 20$ min, the corresponding price is $u(20) = \$4.026$. The optimal throughput of the HOT lanes is 30 veh/min. Comparing the numerical results with the analytical results, we find that our method indeed drive the system to the optimal state as expected.

For the feedback method (3.16), we set $u(0) = \ln 2$, and $K_I = 0.01(\$\cdot\text{min})$ for the controller. The numerical results are shown in Figure 3.7. Since the demand is always higher than the capacity of the HOT lanes, the price increases with time as shown in Figure 3.7b. In Figure 3.7a the queue lengths on both lanes increase. Thus the zero-queue condition on the HOT lanes cannot be guaranteed by the I-controller. This verify our conclusion that the dynamic price determined by a single I-controller cannot achieve the two operation objectives for the HOT lanes at the same time.

(a) Queue lengths   (b) Price
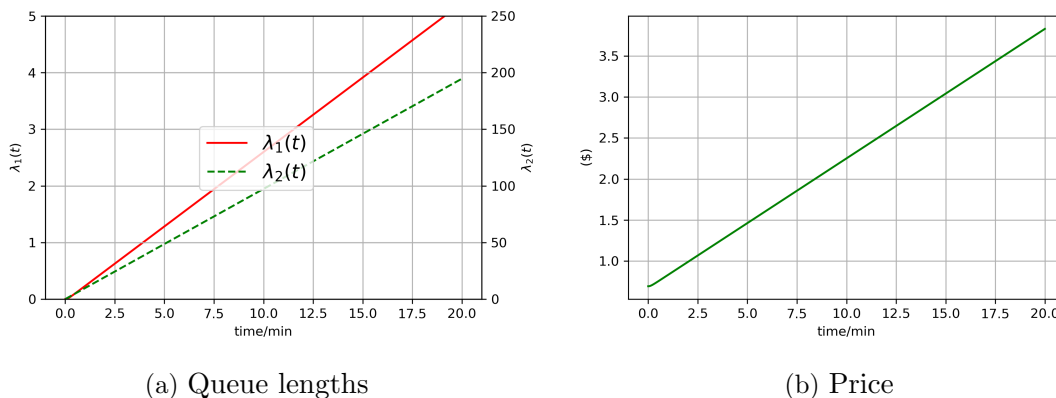
Figure 3.7: Numerical results of the feedback method in (Yin and Lou, 2009).

For the self-learning method (3.17), we set the true value to be $[\alpha_1; \alpha_2; \gamma] = [0.5; 1; 0]$, and the initial guess is $[\alpha_1(0); \alpha_2(0); \gamma(0)] = [0.25; 1; 0.1]$. And the variance of measurement noise is set to be 0.09.
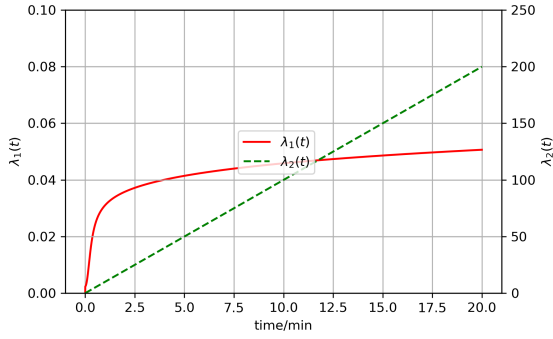
(a) Queue lengths
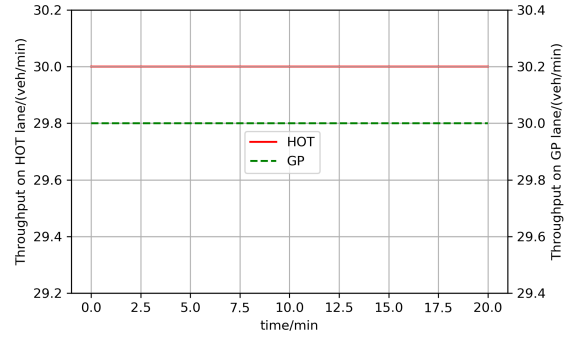
(b) Throughputs
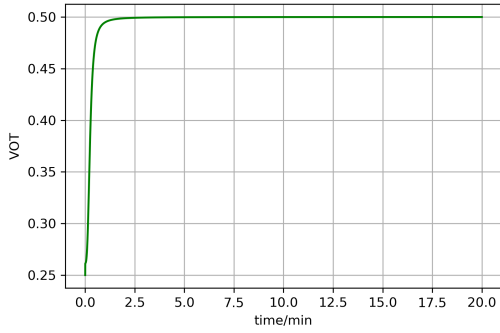
(c) Estimated average VOT
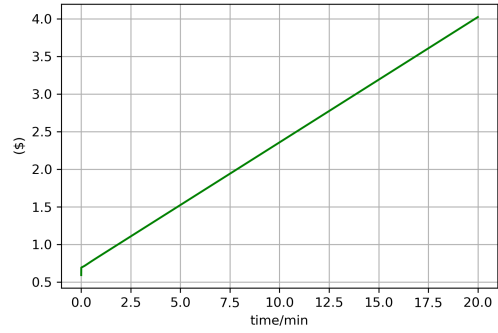
(d) Price

Figure 3.8: Numerical results of the self-learning method in (Yin and Lou, 2009).

Since the demand of paying SOVs is higher than the optimal value, queue length increases on the HOT lanes, as shown in Figure 3.8a. So, the optimal state cannot be guaranteed. Since the demand of paying SOVs is not less than 20 veh/min, the throughput of the HOT lanes is always 30 veh/min. In Figure 3.8c, $\pi(t)$ rises from \$0.25/min, and converges to \$0.5/min. However, different from Figure 3.6c, there is no overshoot in the estimation process. In Figure 3.8d, the dynamic price increases with time, and it is \$4.024 at 20 minutes. We also investigate the impacts of the measurement noise in the KF and the initial guess of α_1 on the length of the residual queue on the HOT lanes. It shows that the queue length will be longer if there is a larger measurement noise or a smaller initial guess of α_1.

Comparing with the system performance of those three methods, we conclude that our method is more effective.

3.7.2   Robustness

In reality, the demand varies with the time, the measurements are inaccurate subject to disturbances, and the model parameters can change with time. Those factors are the main sources of the stochasticity for the HOT lanes system. Robustness is the degree to which a system can function correctly in the presence of uncertainty. Providing robustness is one of the key functions for the feedback control. In this subsection, we examine the robustness of our controller subject to disturbances in the demand pattern. We assume $q_1(t)$ is Poisson with an average of 10 veh/min, and $q_2(t)$ is Poisson with an average of 60 veh/min. For the controller, we keep the same setup, i.e., $K_1 = \$0.1/\text{min}^2$ and $K_2 = \$0.1/\text{min}$.
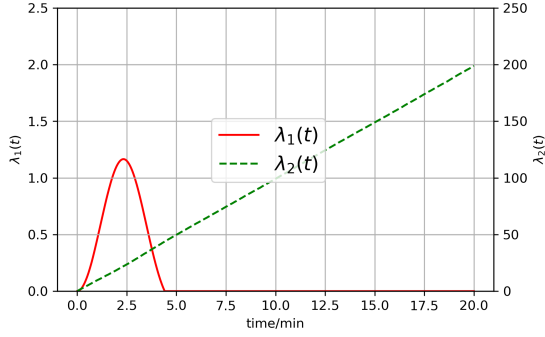
(a) Queue lengths
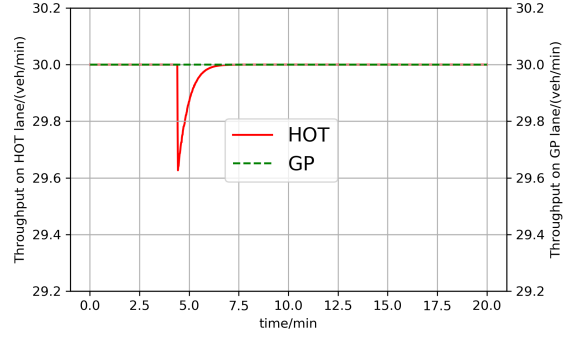
(b) Throughputs
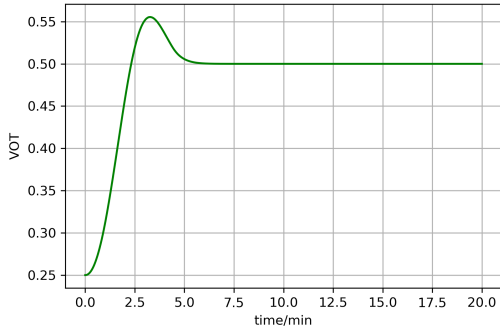
(c) Estimated average VOT
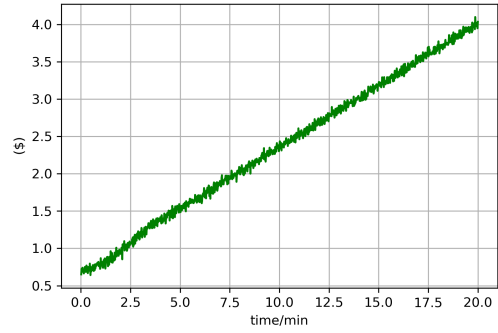
(d) Price

Figure 3.9: Numerical results of our method ((3.19) and (3.20)) with random demands.

As shown in Figure 3.9c, we can still estimate the average VOT with the random demand pattern. Same as the results in Figure 3.6a and 3.6b, no queue exists on the HOT lanes and the throughput of the HOT lanes is at capacity after the system reaches the optimal state (see Figure 3.9a and Figure 3.9b). At the same time, we observe some fluctuations in the price in Figure 3.9d, which is caused by the random demand of HOVs and SOVs. However, comparing with Figure 3.6d, the price is still around the price for the constant demand pattern. In this sense, we conclude that the controller is robust with respect to random variations in the demand patterns.

3.7.3 Stability of the system

In this section, we provide numerical results of the original and the approximate model in Section 3.6.

The original model

In this subsection, we numerically solve (3.21) with different parameters, and subject to different initial queue length to the equilibrium state (different initial queues).

For the first pattern, we set $K_1 = \$0.1/\text{min}^2$, $K_2 = \$0.1/\text{min}$, $\lambda_1(0) = 1$ veh, and $\pi(0) = \$0.25/\text{min}$.

<center>(a)</center>

<center>(b)</center>
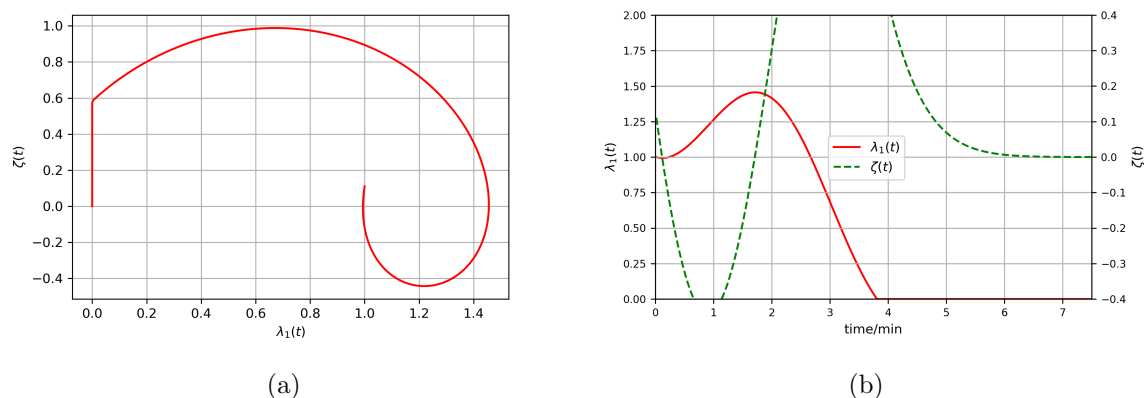
Figure 3.10: Numerical results of the original model (convergence pattern 1 in 3.6.1).

Figure 3.10a is the phase diagram, the horizontal axis represents $\lambda_1(t)$, and the vertical axis represents $\zeta(t)$. The staring point is (1,0.11). Initially, $\zeta(t)$ decreases till reaches -0.44 veh/min, and the maximum $\lambda_1(t)$ is 1.46 veh. It is obvious that $\lambda_1(t)$ reaches 0 earlier than $\zeta(t)$ in both figures. Figure 3.10b shows how $\lambda_1(t)$ and $\zeta(t)$ change with the time. The horizontal axis is the time, the left vertical axis represents $\lambda_1(t)$ and the right vertical axis represents $\zeta(t)$. $\lambda_1(t)$ reaches 0 at around 4 minutes, and after that $\zeta(t)$ converges to 0 in a

<center>49</center>

Gaussian manner.

To obtain the second convergence pattern, we set $K_1 = \$0.1/\text{min}^2$, $K_2 = \$0.2/\text{min}$. The initial condition is the same as the first pattern.

(a)   (b)
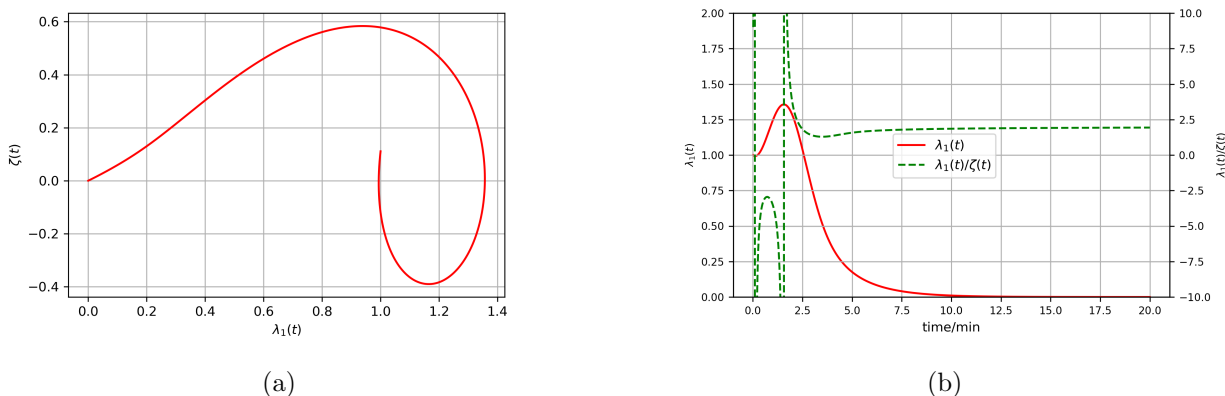
Figure 3.11: Numerical results of the original model (convergence pattern 2 in 3.6.1).

In Figure 3.11a, the starting point is $(1, 0.11)$. The minimum value of $\zeta(t)$ is -0.39 veh/min, and the maximum value of $\lambda_1(t)$ is 1.36 veh. Both $\lambda_1(t)$ and $\zeta(t)$ converges to 0 exponentially after a long time. In Figure 3.11b, the right vertical axis represents the ratio between $\lambda_1(t)$ and $\zeta(t)$. which is different from Figure 3.10b. $\lambda_1(t)$ converges exponentially staring from about 2 minutes, and reaches 0 at around 20 minutes. $\lambda_1(t)/\zeta(t)$ converges to 2, which equals K_2/K_1.

We further numerically check the convergence pattern. With $\lambda_1(0) = 1$ veh and $\pi(0) = \$0.25$ /min, when we set $K_1 = \$0.1/\text{min}^2$, the two phases switch around $K_2 = \$0.14/\text{min}$: a larger K_2 leads to congested, exponential convergence; and a smaller K_2 leads to uncongested, Gaussian convergence (Theorem 3.6.1).

When the demands are stochastic, we can observe similar convergence patterns as Figure 3.10 and 3.11.

The approximate model

In this subsection, we numerically solve the approximate model of (3.21a) and (3.23).

For the first pattern, we set $K_1 = \$0.1/\text{min}^2$, $K_2 = \$0.1/\text{min}$, $\lambda_1(0) = 1$ veh, and $\zeta(0) = 0.11$ veh/min. Note that $\zeta(0) = 0.11$ veh/min is chosen based on the starting point calculated from the original model.

(a)   (b)
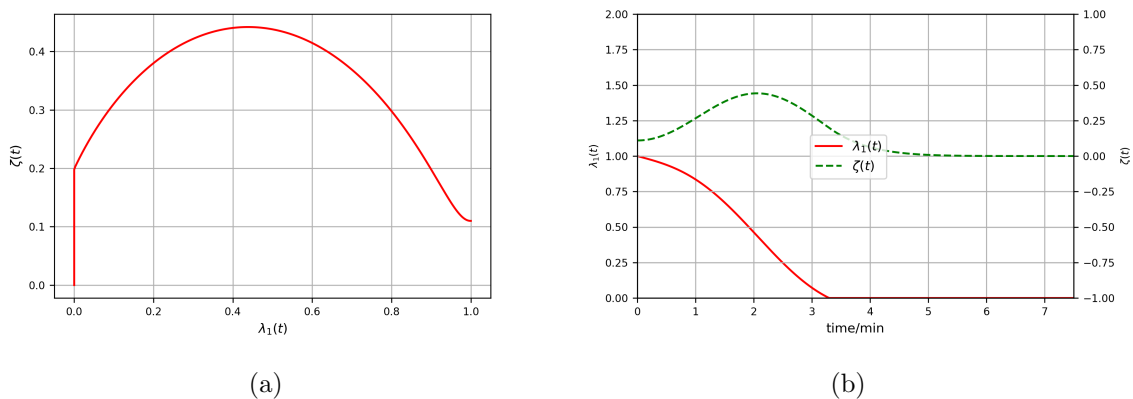
Figure 3.12: Numerical results of the approximate model (convergence pattern 1 in 3.6.1).

In Figure 3.12a, the trajectory starts from (1,0.11), and ends at (0,0). The maximum value of $\zeta(t)$ is 0.44 veh/min. The convergence pattern is different from Figure 3.10a initially, because we drop the first term on the left-hand side of (3.22) when deriving (3.23). After a long time, the convergence pattern is similar to Figure 3.10a. From Figure 3.12b, we can observe that $\lambda_1(t)$ decreases until it reaches 0, while $\zeta(t)$ increases to 0.44 and then converges to 0 in a Gaussian manner.

For the second pattern, we set $K_1 = \$0.1/\text{min}^2$, $K_2 = \$0.2/\text{min}$, $\lambda_1(0) = 1$ veh, and $\zeta(0) = 0.11$ veh/min.

(a)   (b)
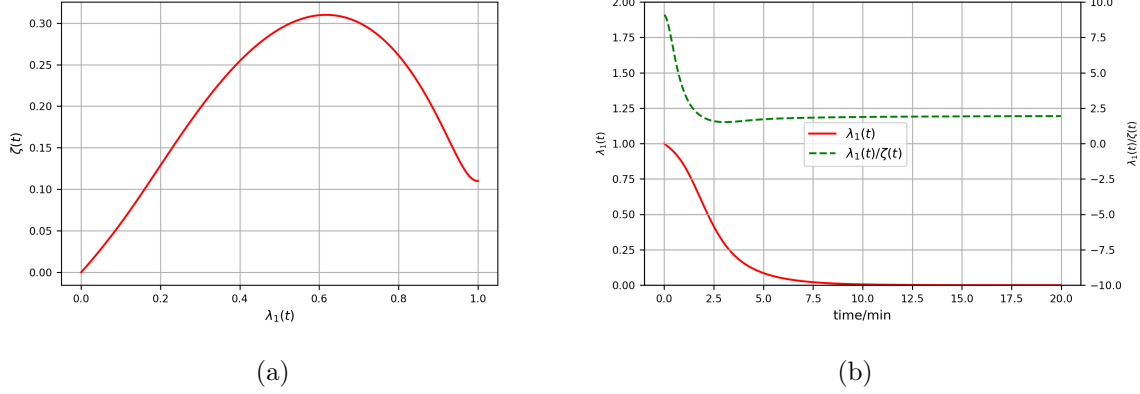
Figure 3.13: Numerical results of the approximate model (convergence pattern 2 in 3.6.1).

In Figure 3.13a, the convergence pattern is similar to Figure 3.11a after a long time. The starting point is (1,0.11), and ends at (0,0). The maximum value of $\zeta(t)$ is 0.31 veh/min. In Figure 3.13b, $\lambda_1(t)$ converges to 0 exponentially (Theorem 3.6.1); and $\lambda_1(t)/\zeta(t)$ converges to 2 after a long time, which is the same as Figure 3.11b.

We also numerically check the convergence pattern. With $\lambda_1(0) = 1$ veh and $\zeta(0) = 0.11$ veh/min, when we set $K_1 = \$0.1/\text{min}^2$, the two phases switch around $K_2 = \$0.14/\text{min}$. Same as the original model, a larger K_2 leads to congested, exponential convergence; and a smaller K_2 leads to uncongested, Gaussian convergence (Theorem 3.6.1).

3.7.4   The scale parameter

In the previous analysis, we assume the operators know the exact value of the scale parameter. In this section, we relax this assumption. Here, the operators make a initial guess of the scale parameter, denoted as α, when determining the dynamic price. Then, the price is set as

$$u(t) = \pi(t)w(t) + \frac{\ln \frac{q_1 + q_2 - C_1}{C_1 - q_1}}{\alpha}. \tag{3.26}$$

52

The setup is the same as Section 3.7.1, except that $\alpha = 1.2$.
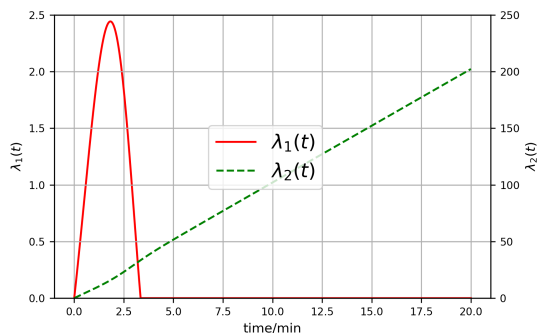
(a) Queue lengths
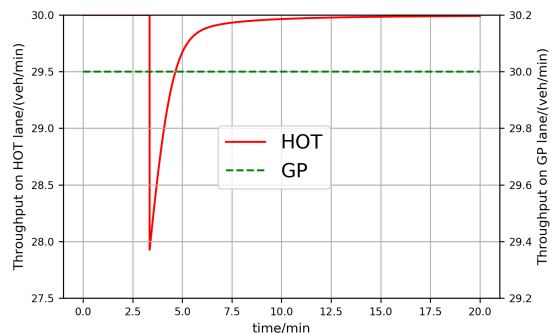
(b) Throughputs
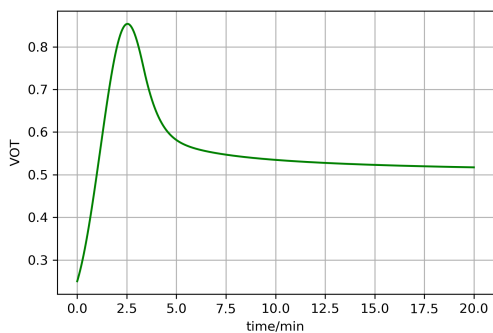
(c) Estimated average VOT
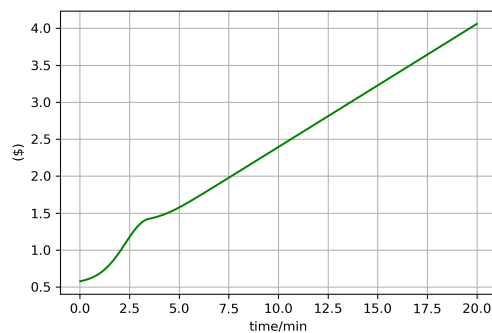
(d) Price

Figure 3.14: Numerical results of our method ((3.19) and (3.26)) with $K_1 = \$0.1/\text{min}^2$ and $K_2 = \$0.1/\text{min}$.

In Figure 3.14a, no queue exists on the HOT lanes after some time, which is the consistent with Figure 3.6a. The throughputs on both lanes are 30 vph eventually (see Figure 3.14b). So, the price in (3.26) drives the system to the optimal state. In Figure 3.14c, the estimated average VOT is approaching the true value. The price is $4.061 at 20 minutes in Figure 3.14d, which is slightly lower than the one in Figure 3.6d.

Then, we test the impact of the scale parameter on the VOT estimation after a long time. The study duration is 1500 minutes. As shown in Figure 3.15, $\pi(t)$ converges to π^*. The result is consistent with the statement that the VOT is not affected by the scale parameter

(Train, 2009).
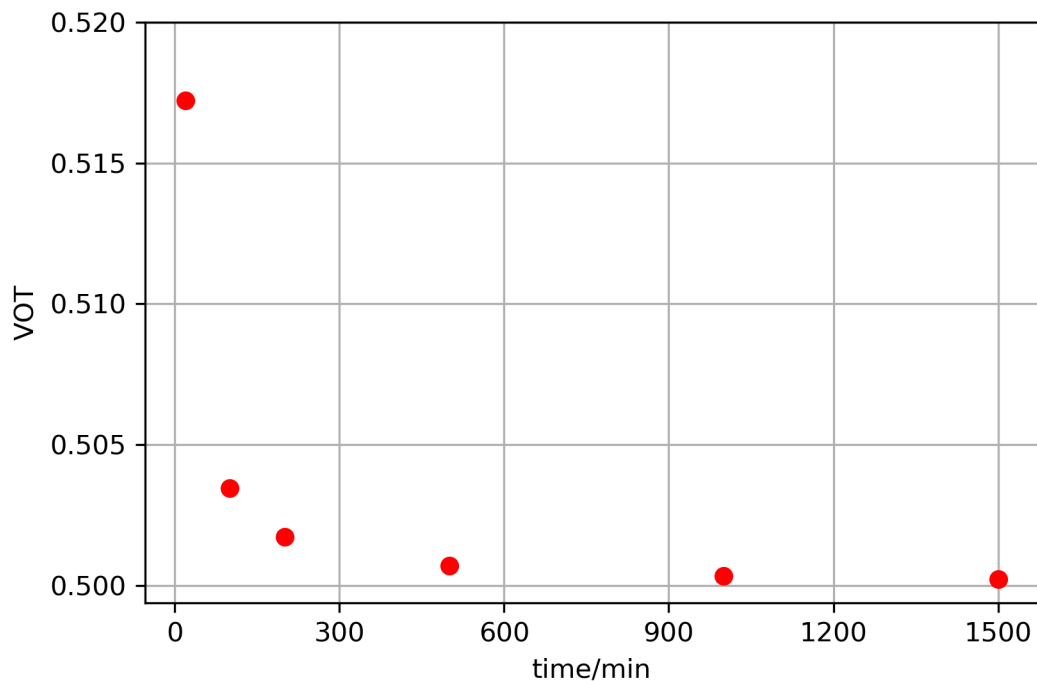
Figure 3.15: The estimated average VOT.

3.8 Conclusion

In this study, we provide a new control theoretic approach to solve the simultaneous estimation and control problem for a traffic system with HOT lanes. We define a new variable called the residual capacity for the HOT lanes. We apply two PQMs to describe the traffic dynamics on the HOT and GP lanes, and capture the lane choice of SOVs with a logit model. The controller can estimate the average VOT and calculate the dynamic price simultaneously, which has received little attention in the field of traffic control. We also analytically prove that the closed-loop system is stable and guaranteed to converge to the optimal state. We provide numerical examples to show that our method is effective and robust respect to randomness in the demand pattern. We find the system converges to the equilibrium state

in the patterns predicted by Theorem 3.6.1. At the same time, we show that the two methods in (Yin and Lou, 2009) are either unstable or cannot guarantee the convergence to the optimal state, and our controller is more efficient and leads to a better system performance. A simple analytical example with constant demand gives us an guideline for the controller design: when the HOT lanes are underutilized while the whole system is congested, the price for the HOT lanes should increase with time. Then, we design a simple I-controller to estimate the average VOT of SOVs, since the average VOT is constant. We also provide a novel analytical proof of the stability to fill the gap in theoretic studies on the HOT lanes. At last, we numerically show that the scale parameter does not affect the estimation of VOT and the optimal state after a long time.

In summary, this study makes three fundamental contributions: (1) to present a simpler formulation of the point queue model based on a new concept of residual capacity, (2) to propose a simple feedback control theoretic approach to estimate the average value of time and calculate the dynamic price based on the logit model, and (3) to analytically and numerically prove that the closed-loop system is stable and guaranteed to converge to the optimal state, in either Gaussian or exponential manners.

The following are some potential future research topics.

- We are interested in the estimation and control problem in a more complex network. First, we will consider the effect of lane changing (Jin, 2013) and capacity drop (Jin et al., 2015) at one freeway segment. Besides the PQM, we will consider the application of the cell transmission model (CTM) to model the traffic dynamics (Daganzo, 1994). Then, we want to look into a more complex scenario with multiple entrances and multiple exits (Pandey and Boyles, 2018, 2019). We have a desire to explore the pricing scheme for a multi-segment HOT system. Intuitively, if a paying SOV wish to occupy the HOT lanes for multiple segments, the price s/he pays should not be a

simple summation of prices for each individual segment.

- In this study, we capture SOVs' lane choice by a logit model with the average VOT. In the future, we want to introduce a new choice model called "vehicle-based" user equilibrium, which is provided in the next chapter. The basic idea is as follows: if a single SOV chooses a lane (HOT or GP), then the cost of this lane is less than or equal to that of the other lane (GP or HOT) which would be experienced by the same vehicle. With the vehicle-based UE principle, we can consider the heterogeneous VOTs of different SOVs.

- Another research topic would be a simultaneous departure time and lane choice model for SOVs (Boyles et al., 2015). For SOVs, their travel costs include a free-flow travel time, a queuing time, a schedule delay and a dynamic price. We want to design pricing schemes considering departure time user equilibrium.

Chapter 4

A Stable Dynamic Pricing Scheme Independent of Lane Choice Models for High-Occupancy Toll Lanes

4.1 Introduction

In this study we consider the control and estimation problems of the same traffic system as in (Yin and Lou, 2009; Lou et al., 2011; Wang and Jin, 2017). We want to determine dynamic prices that: (1) keep zero queue on the HOT lanes; and (2) maximize the HOT lanes' throughput. The main contributions of this study are: (1) proposing a general lane choice for SOVs based on the travel time difference, the price, and the SOVs' social and economic characteristics; (2) designing a feedback control method to directly calculate dynamic prices for the HOT lanes, based on the real-time traffic states; (3) estimating the value of times (VOTs) if a specific lane choice model is provided.

The rest of the study is organized as follows. In Section 4.2, we describe the traffic system

with HOT lanes, including the point queue models and provide general information of the lane choice models. In Section 4.3, we explore some properties of the general lane choice model, and present two specific models: a logit model and a vehicle-based user equilibrium principle. In Section 4.4, we present a feedback control strategy to determine dynamic prices and study the properties of the closed-loop system. In Section 4.5, we show the process for estimating the distribution of VOTs. In Section 4.6, we numerically study the control and estimation problem considering different lane choice models. In Section 4.7, we conclude the study with discussions on future research topics.

4.2   System description and problem statement

The traffic system we are going to analyze is the same as Section 3.2, and the same point queue model (PQM) is applied to capture the traffic dynamics. Different from the previous study, we use a general choice model to describe SOVs' lane choice. We define the following variables to capture the traffic dynamics:

- Arrival rates: $q_1(t)$ and $q_2(t)$ are the arrival flow-rates of HOVs and SOVs, respectively, at time t.

- Capacities: C_1 and C_2 are the capacities of the HOT and GP lanes, respectively.

- Queue lengths: $\lambda_1(t)$ and $\lambda_2(t)$ are the queue lengths on the HOT and GP lanes, respectively.

- Waiting times: $w_1(t)$ and $w_2(t)$ are the waiting times for vehicles leaving from the downstream end at t on the HOT and GP lanes, respectively.

Same as the previous study, there are two assumptions regarding the demand patterns: (1)

the system is congested during the study period $t \in [0, T]$; i.e.,

$$\int_0^t q_1(s) + q_2(s)ds > (C_1 + C_2)t.$$

(2) HOVs can use the HOT lanes for free, but the demand (arrival rate) of HOVs is below the HOT lanes' capacity; i.e.,

$$q_1(t) < C_1.$$

We further define the following variables:

- The price: $u(t)$ is the time-dependent price paid by SOVs who use the HOT lanes, not the GP lanes.

- The flow-rate of paid SOVs: $q_3(t)$ is the flow-rate of paid SOVs, which are SOVs but use the HOT lane.

- The proportion of SOVs choosing the HOT lanes: $p(t)$,

$$p(t) = \frac{q_3(t)}{q_2(t)} \in [0, 1]. \tag{4.1}$$

4.2.1 Point queue models

For the point queues on the two lane groups, the waiting times are determined by the arrival rates and capacities. Their arrival rates are respectively $q_1(t) + q_3(t)$ and $q_2(t) - q_3(t)$. Thus the queue changing rates on both lanes are

$$\frac{\mathrm{d}}{\mathrm{d}t}\lambda_1(t) = \max\{q_1(t) + q_3(t) - C_1, -\frac{\lambda_1(t)}{\epsilon}\}, \tag{4.2a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\lambda_2(t) = \max\{q_2(t) - q_3(t) - C_2, -\frac{\lambda_2(t)}{\epsilon}\}, \tag{4.2b}$$

59

where $\epsilon = \lim_{\Delta t \to 0^+} \Delta t$ is an infinitesimal number and equals Δt in the discrete form, and

$$\zeta(t) = C_1 - q_1(t) - q_3(t) \tag{4.3}$$

is the residual capacity of the HOT lanes. The waiting times on two types of lanes are given by

$$w_1(t) = \frac{\lambda_1(t)}{C_1}, \tag{4.4a}$$

$$w_2(t) = \frac{\lambda_2(t)}{C_2}. \tag{4.4b}$$

The additional waiting time on the GP lanes is given by

$$w(t) = w_2(t) - w_1(t) = \frac{\lambda_2(t)}{C_2} - \frac{\lambda_1(t)}{C_1}. \tag{4.5}$$

4.2.2 Lane choice model

Intuitively, the proportion of SOVs choosing the HOT lanes depends on the difference in the waiting times on the two types of lanes, $w(t)$, the price, $u(t)$, and the SOVs' social and economic characteristics. Here we represent the general lane choice model by the following functional relationship:

$$Pr(t) = G(w(t), u(t)), \tag{4.6}$$

where the functional form $G(\cdot, \cdot)$ depends on the SOVs' characteristics, such as their VOTs. From the definition of $Pr(t)$ and $\zeta(t)$ in (4.1) and (4.3) respectively, (4.6) leads to the

following equation of $\zeta(t)$:

$$\zeta(t) = C_1 - q_1(t) - q_2(t)G(w(t), u(t)). \tag{4.7}$$

Here we assume that $Pr(t)$ does not depend on the historical values of $w(t)$ or $u(t)$. We also ignore the psychological effects of pricing[1]. We also assume that the drivers are fully aware of $w(t)$ and $u(t)$, which can help to mitigate the psychological impacts.

We will provide more details about the lane choice model, and present two models that satisfy those requirements in Section 4.3.

4.2.3 Two problems

For the traffic system with HOT lanes, there are two outstanding problems. The first is to determine a dynamic price, $u(t)$, for time-dependent arrival rates of both HOVs and SOVs; and the second is to estimate the distribution of SOVs' VOTs, $f(\pi)$.

This study is different from the literature in the following aspects: (1) in (Yin and Lou, 2009; Lou et al., 2011; Wang and Jin, 2017), a logit model was assumed for the lane choice. Here we solve the control problem first without relying on knowledge of travelers' underlying lane choice behaviors. In addition, we solve the estimation problem with the logit model with an average VOT as well as the vehicle-based user equilibrium principle with heterogeneous VOTs; and (2) in (Gardner et al., 2013), the dynamic prices are directly calculated from given distribution of VOTs or average VOT. However, in this study, the dynamic prices are calculated without knowing the lane choice model or its parameters.

Note that the traffic and lane choice models are used to describe the dynamics of the traffic

---

[1]Drivers consider the price as an indication of time savings and congestion. Psychological effects of pricing state that higher prices would provide greater time savings (Janson and Levinson, 2014).

system, but not used in the design of the controller. Thus the controller is independent of these models and could apply to other mathematical models, simulators, or real-world systems.

4.3 Lane choice models

Since the psychological effects of pricing is ignored, ideally, the lane-choice model, (4.6), satisfies the following conditions: (with $u \geq 0$ and $w \geq 0$).

1. The choice proportion increases in $w(t)$; i.e., the larger the difference in waiting times, the more SOVs will choose to pay for the HOT lanes:

$$\frac{\partial G(w, u)}{\partial w} > 0.$$

2. The choice proportion decreases in $u(t)$; i.e., the higher the price, the fewer SOVs will choose to pay for the HOT lanes:

$$\frac{\partial G(w, u)}{\partial u} < 0.$$

3. Since the choice proportion $Pr(t) \in [0, 1]$, the function $G(\cdot, \cdot)$ should also satisfy the following conditions: $G(\infty, u) = 1$ for a constant u; $G(w, \infty) = 0$ for a constant w; $G(0, u) = 0$ for $u > 0$; $G(w, 0) = 1$ for $w > 0$; $G(0, 0) = 0$.

From the respective definitions of $p(t)$ and $\zeta(t)$ in (4.1) and (4.3), the general lane-choice model, (4.7), leads to the following relation between the residual capacity and the price:

$$u(t) = \Phi(\zeta(t); w(t)) \equiv G^{-1}(w(t), p(t)) = G^{-1}\left(w(t), \frac{C_1 - q_1(t) - \zeta(t)}{q_2(t)}\right). \tag{4.8}$$

Note that the above function is determined by SOVs' lane choice behaviors, which could vary with respect to time, location, and combination of vehicles. Therefore, the above function is unknown to the system operators.

In the following section, we present two examples with a logit model and a vehicle-based user equilibrium principle.

4.3.1 A logit model

If SOVs choose the HOT lanes based on the logit model, then (4.6) is replaced by

$$Pr(t) = \frac{1}{1 + \exp(\alpha^*(u(t) - \pi^* w(t)))}. \tag{4.9}$$

Here the measurable utility for paying SOVs is $-(u(t) + \pi^*(t_f + w_1(t)))$, and that for non-paying SOVs is $-\pi^*(t_f + w_2(t))$, where t_f is the free-flow travel time. π^* represents the true average VOT, and it is unknown to the operators. For simplicity, we assume $\alpha^* = 1$. Then (4.8) in this case can be written as

$$u(t) = \pi^* w(t) + \ln \frac{q_1(t) + q_2(t) - C_1 + \zeta(t)}{C_1 - q_1(t) - \zeta(t)}. \tag{4.10}$$

4.3.2 Heterogeneous values of time and vehicle-based user equilibrium principle

We assume that different SOVs have heterogeneous values of time (VOTs). For vehicle i, the VOT is denoted by π_i. We assume that the VOT follows a distribution, $f(\pi)$, which could be discrete or continuous or mixed. In general, the probability distribution function is unknown to the system operator.

In the original Wardrop's first (user equilibrium, UE) principle, "The journey times on all the routes actually used are equal, and less than (or equal to) those which would be experienced by a single vehicle on any unused route". Here we extend the UE principle for individual vehicles choosing different lanes based on travel costs. In the following vehicle-based UE principle, if a single SOV chooses a lane (HOT or GP), then the cost on the lane is less than or equal to that on the other lane (GP or HOT) which would be experienced by the same vehicle. If we replace the costs by the journey times and the lanes by the routes, then the vehicle-based UE principle is equivalent to the original Wardrop's UE principle. The vehicle-based UE principle can be easily applied with explicit VOTs and congestion pricing. We can see that the vehicle-based UE principle is consistent with the selfish routing principle.

Therefore, if SOV i chooses the HOT lanes at t, then

$$w_1(t)\pi_i + u(t) \leq w_2(t)\pi_i; \tag{4.11a}$$

if SOV j chooses the GP lanes, then

$$w_2(t)\pi_j \leq w_1(t)\pi_j + u(t). \tag{4.11b}$$

Assuming that π_i (π_j) is a continuous random variable and $w_2(t) > w_1(t)$, we have the following conclusion.

Lemma 4.3.1. *At t, the proportion of SOVs choosing the HOT lanes is given by*

$$Pr(t) = 1 - F\left(\frac{u(t)}{w(t)}\right), \tag{4.12}$$

where $F(\cdot)$ is the cumulative distribution function of $f(\cdot)$.

Proof. From (4.11a), we can see that, for any SOV i choosing the HOT lanes,

$$\pi_i \geq \frac{u(t)}{w(t)};$$

for any SOV j choosing the GP lanes,

$$\pi_j \leq \frac{u(t)}{w(t)}.$$

Therefore, the proportion of SOVs choosing the HOT lanes is given by (4.12). $\qquad\square$

Note that (4.12) was first presented in (Gardner et al., 2013), but without relating it to the vehicle-based UE principle.

Correspondingly, (4.8) in this case can be written as

$$u(t) = z\left(\frac{C_1 - q_1(t) - \zeta(t)}{q_2(t)}\right)w(t), \tag{4.13}$$

where $z(Pr)$ is $100(1 - Pr)$th-percentile, defined by

$$Pr(t) = 1 - F(z(Pr(t))).$$

As an example, if the VOTs follow an exponential distribution, $F(x) = 1 - \exp(-x/\pi^*)$, where π^* is the average VOT, then we have

$$u(t) = \pi^* w(t) \ln \frac{q_2(t)}{C_1 - q_1(t) - \zeta(t)}. \tag{4.14}$$

If the VOTs follow the Burr distribution in (Gardner et al., 2013), then we have

$$u(t) = \pi^* w(t) \left(\frac{q_1(t) + q_2(t) - C_1 + \zeta(t)}{C_1 - q_1(t) - \zeta(t)}\right)^{1/\gamma}, \tag{4.15}$$

where γ is a shape parameter. Based on the U.S. income distribution in 2008, the 25th percentile to median ratio shows $\gamma \approx 1.5$, and the 75th percentile to median ratio gives $\gamma \approx 2$ (Gardner et al., 2013).

4.3.3 A general lane choice model

For both the logit and UE models of lane choices, the relationship between the price and the residual capacity can be written in the following form:

$$u(t) = A(\zeta(t))w(t) + B(\zeta(t)), \tag{4.16}$$

Here $A(\zeta(t)) > 0$, $A'(\zeta) = \frac{\mathrm{d}A(\zeta)}{\mathrm{d}\zeta} \geq 0$, $B'(\zeta) = \frac{\mathrm{d}B(\zeta)}{\mathrm{d}\zeta} \geq 0$, and $A'(0) + B'(0) > 0$. Therefore, the price linearly increases in the extra queuing times on the GP lanes, when the two operation objectives are reached. Note that, however, $A(\zeta(t))$ and $B(\zeta(t))$ depend on the lane choice models and their corresponding parameters, which again are both assumed to be unknown to the system operators.

Therefore, the plant dynamics can be modeled by the point queue models for the queuing dynamics on both lanes in (4.2), the definition of the extra waiting time on the GP lanes in (4.5), and the general lane choice model in (4.16), as illustrated in Figure 4.1. Inside the plant, the inputs for the lane choice model are $u(t)$, $q_2(t)$ and $w(t)$, and the output is $\zeta(t)$; as for the PQM, the inputs are $q_1(t)$, $q_2(t)$ and $\zeta(t)$, and the outputs are $\lambda_1(t)$ and $\lambda_2(t)$.
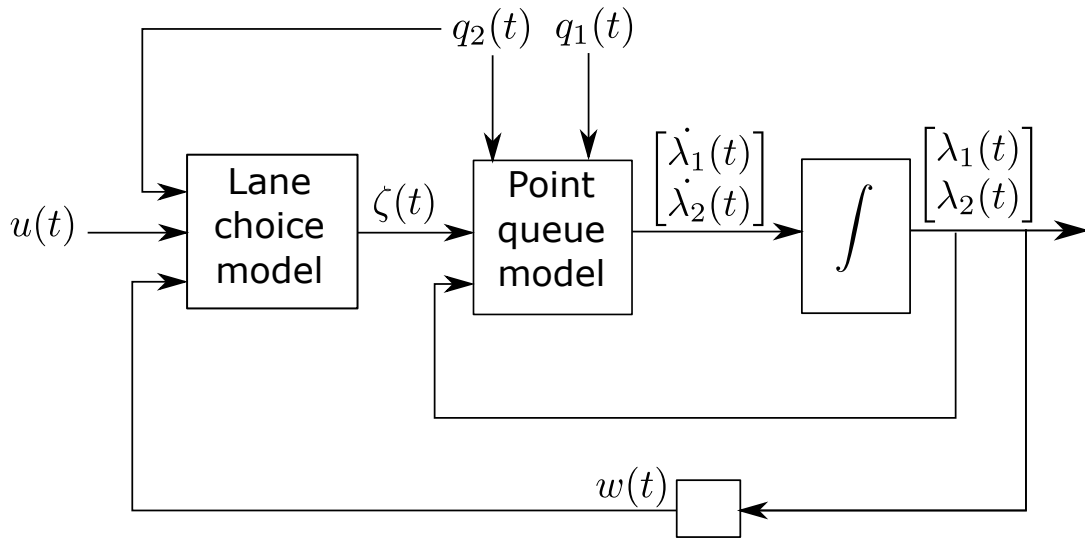
Figure 4.1: Block diagram of the plant.

4.4 A feedback control method

In this section we solve the control problem. Generally, there are two objectives for the control problem (Perez and Sciara, 2003): (1)keeping zero queue on the HOT lanes; and (2) maximizing the HOT lanes' throughput. They serve as two reference points for the system.

1. For the point queue models, the second control objective is equivalent to $\lambda_1(t) = 0$.

2. From the second control objective, the optimal demand of paying SOVs is

   $$q_3(t) = C_1 - q_1(t);$$

   equivalently, the residual queue is zero

   $$\zeta(t) = 0. \tag{4.17}$$

In this study, we desire to set appropriate prices, $u(t)$, to reach such optimal state.

From the general lane choice model in (4.16), we can see that the price should be linearly increasing in the queuing time difference at the optimal state. Thus we propose the following controller

$$u(t) = a(t)w(t) + b(t), \tag{4.18}$$

where $a(t)$ and $b(t)$ are determined by the following I-controllers:

$$\frac{d}{dt}a(t) = K_1\lambda_1(t) - K_2\zeta(t), \tag{4.19a}$$
$$\frac{d}{dt}b(t) = K_3\lambda_1(t) - K_4\zeta(t). \tag{4.19b}$$

Here the units of $a(t)$ and $b(t)$ are $\$/min$ and $\$$ respectively, and those of K_1, K_2, K_3, and K_4 are $\$/veh/min^2$, $\$/veh/min$, $\$/veh/min$, and $\$/veh$, respectively.

If there is a queue on the HOT lanes; i.e., if $\lambda_1(t) > 0$, we increase the price so that fewer SOVs choose the HOT lanes. If $q_3(t) < C_1 - q_1(t)$ or equivalently if $\zeta(t) > 0$, we decrease the price to attract more SOVs to choose the HOT lanes. Therefore, all of the coefficients in (4.19), including K_1, K_2, K_3, and K_4, should be positive.

(4.18) and (4.19) forms a feedback controller, where $w(t)$, $\lambda_1(t)$, and $\zeta(t)$ need to be measured in real time and fed back to the controller to update the dynamic price.

The framework of the control system is shown in Figure 4.2. $q_1(t)$ and $q_2(t)$ are the inputs of the system; $\lambda_1(t)$ and $\lambda_2(t)$ are the state variables; $\lambda_1(t)$, $\lambda_2(t)$, $\zeta(t)$ are the outputs. The plant is consisted of a lane choice model and a traffic model, which has been discussed in Figure 4.1. Dynamic price is calculated by the feedback controller ((4.18) and (4.19)).

Figure 4.2: Block diagram of the control system.

In the following section, we will analyze the equilibrium states and their stability property of the closed-loop system, comprising of (4.2), (4.5), (4.16), (4.18), and (4.19). In addition, we assume that the demands are constant: $q_1(t) = q_1$, and $q_2(t) = q_2$.

4.4.1   Equilibrium states

Combining (4.2a) and (4.19), we obtaining the following system:

$$\frac{d}{dt}\lambda_1(t) = \max\{-\zeta(t), -\frac{\lambda_1(t)}{\epsilon}\}, \tag{4.20a}$$

$$\frac{d}{dt}a(t) = K_1\lambda_1(t) - K_2\zeta(t), \tag{4.20b}$$

$$\frac{d}{dt}b(t) = K_3\lambda_1(t) - K_4\zeta(t). \tag{4.20c}$$

In the equilibrium states, $\dot{\lambda}_1(t) = 0$, $\dot{a}(t) = 0$, and $\dot{b}(t) = 0$, which leads to $\lambda_1(t) = \lambda_1$, $a(t) = a$ and $b(t) = b$. Then from (4.20a) and (4.20b), $\lambda_1(t) = 0$ and $\zeta(t) = 0$. Therefore, the equilibrium states of the closed-loop system are the optimal states.

69

In the equilibrium states, (4.2b) leads to $\dot{\lambda}_2(t) = C_2 w_0$, where $w_0 = \frac{q_1+q_2-C_1-C_2}{C_2}$; thus $\lambda_2(t) = C_2 w_0 t$ [2], and $w(t) = w_0 t$. Both $a(t) = a$ and $b(t) = b$ are constant from (4.19), and the dynamic price in (4.18) is given by

$$u(t) = a w_0 t + b.$$

Further from (4.16) we have

$$u(t) = A(0) w_0 t + B(0).$$

Therefore, we have

$$A(0) = a,$$

$$B(0) = b.$$

That is, $a(t)$ and $b(t)$ are the respective accurate estimators of $A(\zeta(t))$ and $B(\zeta(t))$ at the equilibrium states.

We can see that the dynamic price in equilibrium states linearly increases in time, which behaves similar to an I^2 controller. Note that a constant price determined by a traditional I-controller is not appropriate when the total demand is high. For example, the I-controller proposed in the feedback method in (Yin and Lou, 2009) failed to drive the closed-loop system to the optimal equilibrium states.

---

[2]Here we assume that the initial queue size on the GP lanes, $\lambda_2(0) = 0$, without loss of generality.

4.4.2 Stability property of the equilibrium states

From (4.16) and (4.18) we can eliminate the control variable, $u(t)$:

$$A(\zeta(t)) + \frac{B(\zeta(t))}{w(t)} = a(t) + \frac{b(t)}{w(t)}.$$

Differentiating both sides with respect to t we have

$$\left[A'(\zeta(t)) + \frac{B'(\zeta(t))}{w(t)} \right] \dot{\zeta}(t) = \dot{a}(t) + \frac{\dot{b}(t)}{w(t)} + \frac{B(\zeta(t)) - b(t)}{w^2(t)} \dot{w}(t),$$

where $A'(\cdot)$ and $B'(\cdot)$ are the respective derivatives of $A(\cdot)$ and $B(\cdot)$. Substituting $\dot{a}(t)$ and $\dot{b}(t)$ in (4.19) into the above equation we obtain

$$\left[A'(\zeta(t)) + \frac{B'(\zeta(t))}{w(t)} \right] \dot{\zeta}(t) = K_1 \lambda_1(t) - K_2 \zeta(t) + \frac{K_3 \lambda_1(t) - K_4 \zeta(t)}{w(t)} + \frac{B(\zeta(t)) - b(t)}{w^2(t)} \dot{w}(t).$$

$$(4.21)$$

Here we consider the stability property of the equilibrium states after a long time. When t is very large, $w(t)$ can be approximated by $w_0 t$ near the equilibrium states.

Near the equilibrium states, $\zeta(t)$ is very small, and the left-hand side of (4.21) can be approximated by $[A'(0) + \frac{B'(0)}{w_0 t}]\dot{\zeta}(t)$. Since $K_3 \lambda_1(t) - K_4 \zeta(t)$, $B(\zeta(t)) - b(t)$, and $\dot{w}(t)$ are all bounded near the equilibrium states, the right-hand side of (4.21) can be approximated by $K_1 \lambda_1(t) - K_2 \zeta(t)$. Therefore, near the equilibrium states after a long time, the closed-loop

71

system can be approximated by the following second-order system of $(\lambda_1(t), \zeta(t))$:

$$\dot{\lambda}_1(t) = \max\{-\zeta(t), -\frac{\lambda_1(t)}{\epsilon}\}, \tag{4.22a}$$

$$\dot{\zeta}(t) \approx \frac{1}{A'(0) + \frac{B'(0)}{w_0 t}} \left(K_1 \lambda_1(t) - K_2 \zeta(t)\right). \tag{4.22b}$$

Note that (4.22) is a switching system with two modes (Liberzon, 2003):

- When $\frac{\lambda_1(t)}{\epsilon} \leq \zeta(t)$; i.e., when the queue size on the HOT lanes is very small with $\lambda_1(t) \leq \epsilon\zeta(t)$, (4.22a) leads to $\lambda_1(t) = 0$, and (4.22b) can be simplified as

$$\dot{\zeta}(t) \approx -\frac{K_2}{A'(0) + \frac{B'(0)}{w_0 t}} \zeta(t). \tag{4.23}$$

- Otherwise, when the queue size on the HOT lanes is relatively large; i.e., when $\frac{\lambda_1(t)}{\epsilon} > \zeta(t)$, (4.22) can be simplified as

$$\dot{\lambda}_1(t) = -\zeta(t), \tag{4.24a}$$

$$\dot{\zeta}(t) \approx \frac{1}{A'(0) + \frac{B'(0)}{w_0 t}} (K_1 \lambda_1(t) - K_2 \zeta(t)). \tag{4.24b}$$

Both (4.23) and (4.24) are linear systems. Thus, the feedback controller in (4.18) drives the plant dynamics to an approximate switching linear system after a long time. In spirit, the controller is a feedback linearization approach to nonlinear systems (Khalil, 2002), but our approach is not a standard one.

Theorem 4.4.1. *For positive coefficients K_1 and K_2, the approximate switching linear sys-*

tem, (4.22), is stable.

Proof. In the general lane choice model (4.16), $A'(0) \geq 0$, $B'(0) \geq 0$, and $A'(0) + B'(0) > 0$. Thus $A'(0) + \frac{B'(0)}{w_0 t} > 0$. Then it is straightforward that the eigenvalue of (4.23) is negative, and the real parts of the two eigenvalues of (4.24) are also negative. Thus both linear systems are stable, and the corresponding switching system is stable. $\qquad\square$

4.5 Estimation of values of time

When designing the feedback controller in (4.18) and (4.19), the system operators have no knowledge of drivers' VOTs. However, if the type of lane choice models is given, we can estimate the values of time. In particular, we can estimate the average VOT, π^*, for the logit model, and the probability density function of VOTs, $F(\pi)$, for the UE model.

If the lane choice behaviors are described by the logit model in (4.9), we can estimate π^* as

$$\hat{\pi}^*(t) = \left[u(t) - \ln \frac{q_2(t) - q_3(t)}{q_3(t)} \right] / w(t). \tag{4.25}$$

That is, with the observed $q_2(t)$, $q_3(t)$, and $w(t)$ as well as the imposed price $u(t)$, we can obtain the estimation of π^*.

If the lane choice behaviors are described by the UE model in (4.12), we can estimate the cumulative distribution function of the VOTs as

$$\hat{F}\left(\frac{u(t)}{w(t)}\right) = 1 - \frac{q_3(t)}{q_2(t)}.$$

That is, with the observed $q_2(t)$, $q_3(t)$, and $w(t)$ as well as the imposed price $u(t)$, we can

plot the cumulative distribution function $\hat{F}(\pi) = 1 - \frac{q_3(t)}{q_2(t)}$ against $\pi = \frac{u(t)}{w(t)}$. By taking the derivative of cumulative distribution function to π, we can estimate the probability density function, i.e., $\hat{f}(\pi) = \frac{d\hat{F}(\pi)}{d\pi}$. Note that $q_1(t)$ is not needed for estimating the distribution of the VOTs. But $q_1(t) < C_1$ needs to be satisfied.

4.6   Numerical examples

For the case study, the site is a freeway segment with lane drop downstream of the GP lanes, and the capacity for one HOT and one GP lane is 30 veh/min. The initial queue lengths on the HOT and GP lanes are 1 and 2, respectively ($\lambda_1(0) = 1$ veh and $\lambda_2(0) = 2$ veh). The study period is 20 minutes, and the time-step size is $0.1/60$ min [3]. The demand of HOVs is constant at $q_1(t) = 10$ veh/min, and the demand of SOVs is constant at $q_2(t) = 60$ veh/min.

4.6.1   An example for the logit model

In the following we present an example for the control and estimation problems, assuming that the SOVs choose their lanes based on the logit model. We assume the true average VOT is $\$0.5/min$ ($\pi^* = \$0.5/min$).

As mentioned in Section 4.4, the only requirement for the controller is that all the coefficients in (4.19) are positive. Then we set $K_1 = \$0.1\text{veh}/\text{min}^2$, $K_2 = \$0.1/\text{veh}/\text{min}$, $K_3 = \$0.2/\text{veh}/\text{min}$, and $K_4 = \$0.2 /\text{veh}$. And the initial values are $a(0) = \$0.25 /\text{min}$ and $b(0) = \$0.1$.

In Figure 4.3a, the queue length on the HOT lanes keeps increasing for the first 1 minute, and the maximum queue length is about 2.77 veh. After about 3 minutes, the queue length

---

[3]If we set $\Delta t = 1/60$ min, the system does not converge to the optimal equilibrium state after a long time (when the duration is longer than 250 minutes). However, for smaller Δt, the system always converges to the optimal equilibrium state.

on the HOT lane converges to 0, and the queue length on the GP lane increases with time. When tuning the parameters, we find that if we increase K_1 and K_2 alone, the queue length on the HOT lanes converges to 0 faster; and when we increase K_3 and K_4 alone, there is a smaller maximum queue length on the HOT lanes. In Figure 4.3b, the price increases linearly after the queue on the HOT lanes is eliminated. The price changing rate is about \$0.167 /min when the system reaches the optimal state. From Figure 4.3c, we find that the estimated VOT converges to the true average VOT almost instantaneously. Based on those results, we conclude that the method is effective with the logit lane choice model.

(a) Queue lengths

(b) Price for the HOT lanes

(c) Estimated average VOT

Figure 4.3: Numerical results of our method ((4.18) and (4.19)) with $K_1 = \$0.1$ veh/min^2, $K_2 = \$0.1$ veh/min, $K_3 = \$0.2$ /veh/min, and $K_4 = \$0.2$ /veh with the logit model.

Then, we want to examine the robustness of the designed controller, since providing robust-

ness is one of the key functions of the feedback control. We are interested in how the system behaves when disturbances exist in the demand pattern. We assume $q_1(t)$ is Poisson with an average of 10 veh/min, and $q_2(t)$ is Poisson with an average of 60 veh/min. The parameters in the controllers are $K_1 = \$0.1$ veh/min^2, $K_2 = \$0.1$ /veh/min, $K_3 = \$0.2$ /veh/min, and $K_4 = \$0.2$ /veh.

(a) Queue lengths

(b) Price for the HOT lanes

(c) Estimated average VOT

Figure 4.4: Numerical results of our method ((4.18) and (4.19)) with $K_1 = \$0.1$ veh/min^2, $K_2 = \$0.1$ veh/min, $K_3 = \$0.2$ /veh/min, and $K_4 = \$0.2$ /veh with the logit model (random demand).

Similar to Figure 4.3a, the queue length on the HOT lanes is eliminated after around 3 minutes (see Figure 4.4a). Since the demand pattern is stochastic, the queue length on the HOT is not always 0. Instead, it stays around 0. In Figure 4.4b, the dynamic price is similar to Figure 4.3b, it increases almost linearly with time. The true average VOT can be still be

estimated (see Figure 4.4c). Then we can conclude that the controller is robust respect to random variations in the demand patterns.

In the following, we numerically prove the stability of the system with constant demands. We plot both the phase diagram and the time-series diagram to show the results.

(a)   (b)

Figure 4.5: Stability analysis of system with the logit model.

Figure 4.5a is the phase diagram, the horizontal axis represents $\lambda_1(t)$, and the vertical axis represents $\zeta(t)$. The staring point is (1,-8.62) [4]. Initially, $\zeta(t)$ increases till reaching 2.06 veh/min, and the maximum $\lambda_1(t)$ is 2.77 veh. It is clear that $\lambda_1(t)$ reaches 0 earlier than $\zeta(t)$. Figure 4.5b shows how $\lambda_1(t)$ and $\zeta(t)$ change with time. The horizontal axis is the time, the left vertical axis represents $\lambda_1(t)$ and the right vertical axis represents $\zeta(t)$. $\lambda_1(t)$ reaches 0 at around 3 minutes, and after that $\zeta(t)$ converges to 0. We test different parameters in the controller and different initial conditions, all the results show that the system converges to the optimal state, as proved by (4.22).

---

[4]$\zeta(0) = -8.62 veh/min$ is calculated from the logit model, with known $\lambda_1(0)$, $\lambda_2(0)$, and $u(0)$

4.6.2 An example for the UE principle

In the following, we assume the VOTs follow an exponential distribution, and the average VOT is $0.5/min$. Then, the CDF is expressed by $F(x) = 1 - e^{-2x}$.

For the controller (4.19), we set $K_1 = \$0.1/veh/min^2$, $K_2 = \$0.1 /veh/min$, $K_3 = \$0.2 /veh/min$, and $K_4 = \$0.2 /veh$. And the initial values are $a(0) = \$0.25 /min$ and $b(0) = \$0.1$.

For the same constant demands in Section 4.6.1, the initial price for the HOT lanes is $0.11, which is too high for most SOVs. In this case, almost no SOVs wants to pay and switch to the HOT lanes. Then, a sharp drop in the queue length on the HOT lanes appears at the beginning, as shown in Figure 4.6a. The queue length on the HOT lane becomes zero after some time. Then, the queue on the HOT lanes stays zero, but the queue on the GP lane increases with time. Different values for parameters in the controller are tested as well. When we increase K_1 and K_2 alone, the queue length on the HOT lanes converges faster to the optimal state. In Figure 4.6b, the price increases linearly after the queue on the HOT lanes is eliminated. The price changing rate is about $0.183/min$ when the system reaches the optimal state.

(a) Queue lengths

(b) Price for the HOT lanes

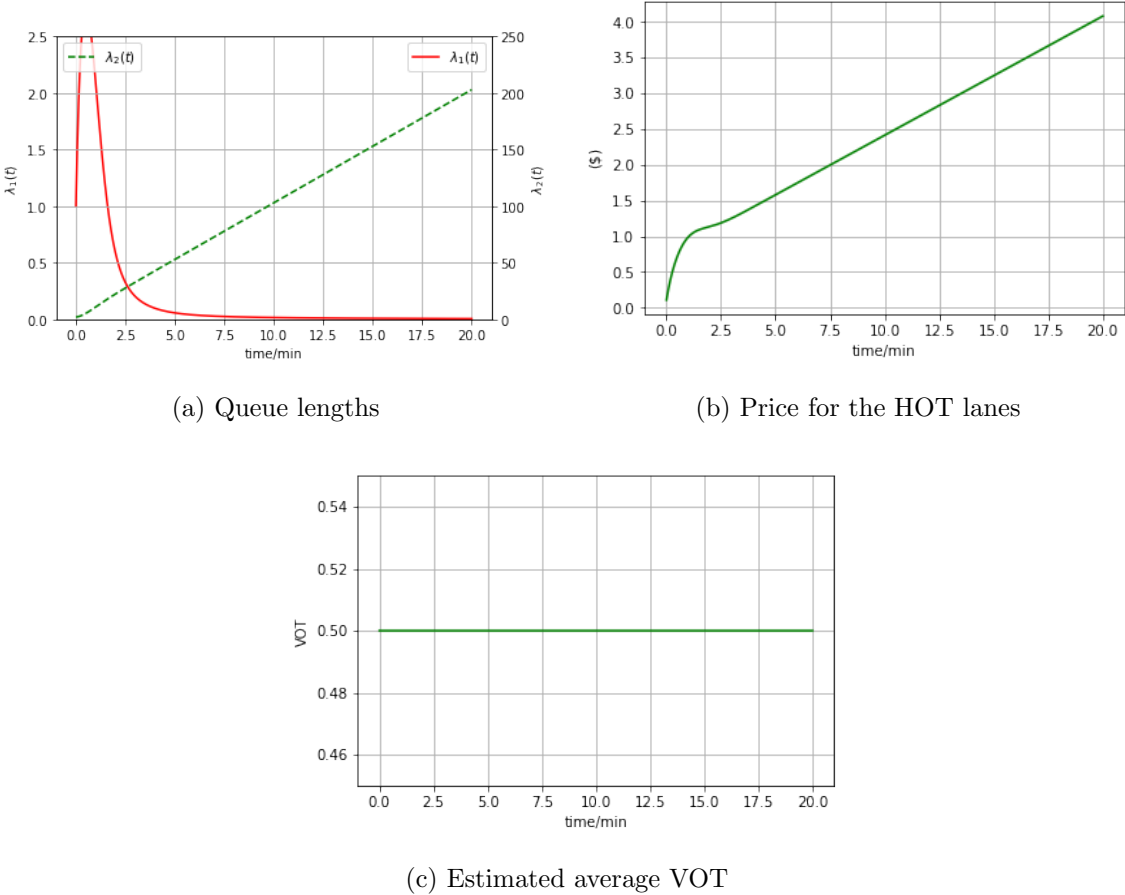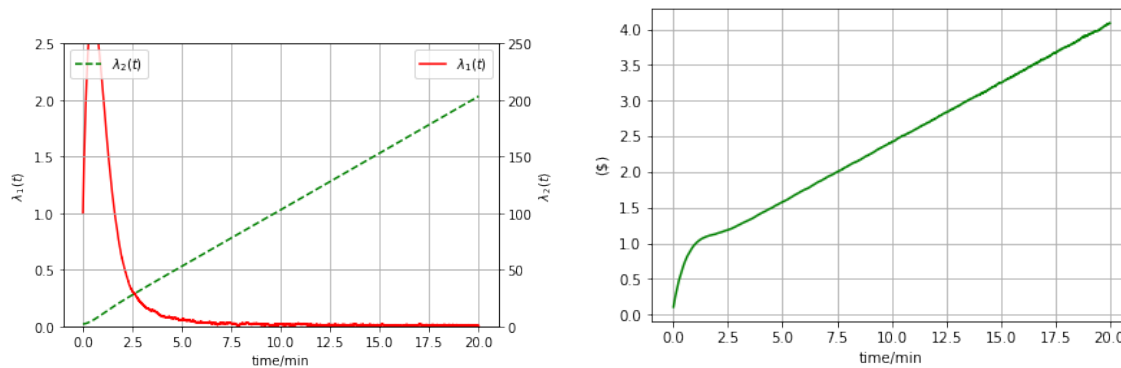Figure 4.6: Numerical results of our method ((4.18) and (4.19)) with $K_1 = \$0.1 \ veh/min^2$, $K_2 = \$0.1 /min$, $K_3 = \$0.2 /veh/min$, and $K_4 = \$0.2 /veh$ with the UE model.

At the same time, we try to estimate the distribution of the VOTs. The CDF is $F(x) = 1 - e^{-2x}$, so it is straightforward that the corresponding probability density function (PDF) is $f(x) = 2e^{-2x}$. We plot the estimated and the theoretic distribution in the figures below.

(a) Estimated vs ground-truth CDF of VOTs   (b) Estimated vs ground-truth PDF of VOTs

Figure 4.7: Estimation of the VOT distribution for the SOVs with the UE model.

In Figure 4.7, the red dots show the estimated CDF and PDF, and the solid lines represent the ground-truth functions. It is obvious that both estimated CDF and PDF are consistent with the ground-truth functions.

We also show the robustness of the controller when the UE model is applied to capture the lane choice. The results in Figure 4.8a and Figure 4.8b are similar to those in Figure 4.6, and the fluctuations are caused by the random demand of HOVs and SOVs. From Figure 4.8c, we can see that although there are disturbances in the demand pattern, we can still estimate the true VOT distribution of SOVs.

(a) Queue lengths

(b) Price for the HOT lanes

(c) Estimated vs ground-truth CDF of VOTs

(d) Estimated vs ground-truth PDF of VOTs

Figure 4.8: System performance and estimation of the VOT distribution with the UE model (random demand).

Then, we numerically study the stability of the system with constant demands. Figure 4.9a is the phase diagram, which shows the relationship between $\lambda_1(t)$ and $\zeta(t)$. The starting point is $(1, 19.91)$, and the system eventually converges to the optimal state $(\lambda^*, \zeta^*) = (0, 0)$. From Figure 4.9b, we can see that $\zeta(t)$ drops near 0 quickly, and then both $\lambda_1(t)$ and $\zeta(t)$ converge to 0 after some time. We also test different parameters in the controller and different initial conditions, the numerical results show that the system converges to the optimal state, as proved by (4.22).

Figure 4.9: Stability analysis of system with the logit model.

Based on the numerical results above, we conclude that our designed controller ((4.18) and (4.19)) can indeed drive the system to the optimal equilibrium state, and it is independent of SOVs' lane choice model. If the type of the lane choice is given, we can estimate the parameters of the model accordingly.

4.7 Conclusion

In this study, we present a new method to solve the control and estimation problem of a traffic system with HOT lanes. We first propose a general lane choice model, and provide examples with a logit model and the vehicle-based user equilibrium principle. Based on the common characteristics of pricing schemes, we obtain an insight regarding the dynamic price: with constant demand, the price should increase linearly when the system reaches the optimal state. Then we develop a novel feedback controller to determine the dynamic prices without knowing SOVs' lane choice models. The equilibrium state and stability properties of the closed-loop are examined analytically and numerically. We also present an estimation method to estimate the parameters of a given lane choice model. We also numerically check the robustness of the controller with stochastic demand patterns.

81

The solution framework is different from those in both groups of studies: (1) different from (Yin and Lou, 2009; Lou et al., 2011; Wang and Jin, 2017), the dynamic prices are first determined by a feedback controller, which does not rely on knowledge of travelers' underlying lane choice behaviors. The controller is designed based on an insight obtained by the analytical solutions to the dynamic pricing problem with constant demands of HOVs and SOVs. The controller is effective independent of the lane choice models and their parameters. However, if the underlying choice model is given to the operator, the VOTs can then be estimated accordingly; and (2) different from (Gardner et al., 2013), the controller is independent of the underlying lane choice models, and the average VOT and the distribution of VOTs are unknown to the operator and will be estimated dynamically.

The following are some potential future research topics:

- We will be interested in studying the control and estimation problems with more realistic traffic flow models. We need to consider the effect of lane changing and capacity drop phenomenon. We also want to explore the simultaneous control and estimation problem for a freeway with multiple entrances and multiple exits.

- We are curious about how reinforcement learning will work for this control and estimation problem. The reinforcement learning algorithm could learn the best toll-rate strategy for all potential traffic states based on real-time or simulated training (Nikolic et al., 2014).

- In the next step, we will consider the third operation objective for the HOT lanes: revenue maximization. We will formulate an optimal control method to explore the dynamic prices that can maximize the revenue for operators.

Chapter 5

Revenue Maximization for High-occupancy Toll Lanes: An Optimal Control Approach

5.1 Introduction

The consistency of the public and private operator's interest is critical in policy implications. Perez and Sciara (2003) claimed that "revenue maximization should generally coincide with the optimization of freeway performances, such as maximizing overall travel-time savings or throughput". But there is no theoretical proof of the statement. Meanwhile, Laval et al. (2015) stated that revenue is maximized for the highest possible price that maintains managed lane at capacity with no queues. However, they make some assumptions: (1) there are only two types of vehicles: paying and non-paying vehicles. They didn't consider the HOVs, which can access the managed lanes for free or at a lower price; (2) all users are identical since they have the same VOT.

Generally, there are three types of toll patterns for HOT lanes: flat rate, time-of-day rate and dynamic rate. Generally, revenue-maximizing toll road operators prefer a dynamic toll because they have an incentive to internalize the congestion costs that users impose on each other (de Palma and Lindsey, 2011). Agnew (1977) formulated an optimal control problem to obtain dynamic prices for a single road that maximize the net benefits of users. He also mentioned that a static model fails to take into account the effect of the toll on the future system loads. Verhoef et al. (1996b) proposed a fixed toll price to maximize profit for a two-link network. Yang and Huang (1997) formulated an optimal control problem to get the dynamic toll that maximizes social benefit of a bottleneck considering departure time choice. However, they didn't consider route choice behavior because only one link exists in this simple network. A revenue-maximizing price model was built for a build–operate–transfer (BOT) toll road equipped with electronic toll collection (Chang and Hsueh, 2006). This paper is one of the first attempts to provide dynamic pricing strategy for revenue maximization on managed lanes. Policy makers started to consider using the revenue collected from toll roads/lanes to cover investment cost after the financial crisis in 2008. Yang (2012) introduced a stochastic control model to maximize the revenue for HOT lanes with multiple entrances and exits. They assumed that the traffic flow switching to the HOT lane had no impact on the traffic condition of GP lanes. However, traffic conditions on different lane groups were proved to be important in setting the tolls. Jang et al. (2014) assumed that the demand of the HOT lane could never be higher than the capacity, and then provided dynamic pricing strategies to achieve different objectives, including minimizing delay and maximizing revenue. The results showed that the revenue-maximization strategy would lead to higher price and less throughput on the HOT lane than the delay-minimization strategy. Göçmen et al. (2015) provided three pricing schemes (non-adaptive time-of-use tolling, myopic pricing and linear adjustment policy) to maximize the revenue on the managed lanes. The myopic policy that individually maximized the expected revenue while keeping the HOT lane in the free-flow condition. They stated that the linear adjustment policy gained the most revenue during

peak hours. Later, Cheng and Ishak (2015) developed a feedback control rule to calculate the dynamic pricing scheme, such that the toll revenue is maximized while the speed on the managed lane is at 45 mph. A logit model was applied to model driver's lane choice, and the VOT was calibrated based on 90 % of drivers' mean hourly income.

In economics, optimal control theory has been applied to study the relationship between pricing schemes and revenue management for decades. Bitran and Caldentey (2003) found a dynamic pricing policy to maximize the expected revenue collected from selling the perishable products over the available period. When the demand for a product is a time-varying linear function of price, Chou and Parlar (2006) formulated an optimal control problem and applied Pontryagins principle to obtain analytical results. In this study, we will apply the optimal control theory to determine pricing strategies to maximize the revenue for a freeway segment with GP and HOT lanes. The rest of this study is organized as follows. In Section 5.2, we provide objective and constraints for the problem. In Section 5.3, we consider a constant demand pattern, then the optimal control problem becomes an optimization problem. We provide analytical solutions when the lane choice for SOVs is modeled by a logit model and a vehicle-based user equilibrium with negative exponentially distributed VOTs. These results can give some insights on the dynamic pricing scheme. In Section 5.4, we formulate an optimal control problem to maximize the revenue for operating the HOT lanes, and propose an iterative algorithm to solve the problem. In Section 5.5, we provide numerical examples with different traffic demand patterns. In Section 5.6, we compare the results of different pricing schemes and discuss policy implications for HOT lanes. In Section 5.7, we conclude this study.

5.2 Problem formulation

We consider a traffic system of a freeway with two types of lanes: the HOT and GP lanes. There are three types of vehicles: the first type is the HOVs, which have two or more passengers; the second type is the regular SOVs, which stay on GP lanes all the time; and the last type is the paid SOVs, which pay a fee to use the HOT lane in order to reduce delay. And the variables are defined as follows:

- Traffic demands: $q_1(t)$ and $q_2(t)$ are demands of HOVs and SOVs, at time t.

- The flow-rate of paid SOVs: $q_3(t)$ is the flow-rate of paid SOVs, which are SOVs but use the HOT lane.

- Capacities: C_1 and C_2 are the capacities of the HOT and GP lanes, respectively.

- Queue lengths: $\lambda_1(t)$ and $\lambda_2(t)$ are the queue lengths on the HOT and GP lanes, respectively.

- Waiting times: $w_1(t)$ and $w_2(t)$ are the waiting times for vehicles leaving at t on the HOT and GP lanes, respectively; $w(t)$ is the waiting time difference on the HOT and GP lanes.

- The proportion of SOVs choosing the HOT lanes: $Pr(t)$,

$$Pr(t) = \frac{q_3(t)}{q_2(t)} \in [0, 1].$$

- The price: $u(t)$ is the time-dependent price paid by SOVs who use the HOT lanes.

We assume the traffic system is congested during the study period $t \in [0, T]$, i.e.,

$$\int_0^T q_1(t) + q_2(t)dt > (C_1 + C_2)T.$$

We also assume the arrival flow-rate of HOVs is below the HOT lane's capacity, i.e., $q_1(t) < C_1$. In this case, some SOVs can pay to use the HOT lane. The flow-rates downstream on GP and HOT lanes should be $q_2(t) - q_3(t)$ and $q_1(t) + q_3(t)$, respectively.

5.2.1 Operation objective

For this problem, we choose $q_3(t)$ as the state variable, and $u(t)$ as the control variable. The objective of this problem is to maximize the revenue of HOT lanes. At any time t, the revenue is the product of paid SOVs and the dynamic price. So, for a time period T, the objective is written as:

$$\max R = \int_0^T q_3(t)u(t)dt. \tag{5.1}$$

5.2.2 Constraints

There are two constraints in the control problem. The first constraint is the traffic dynamics, and the second one is the lane group choice of SOVs.

Traffic flow model

To model traffic dynamics, we apply the point queue model (Vickrey, 1969; Jin, 2015). For HOT and GP lanes in a homogeneous road segment, the free-flow travel time is the same, the only travel time difference exists in the queuing time. The dynamics of the two point

queues are described by the following ordinary differential equations:

$$\frac{d}{dt}\lambda_1(t) = \max\{q_1(t) + q_3(t) - C_1, -\frac{\lambda_1(t)}{\epsilon}\}, \tag{5.2a}$$

$$\frac{d}{dt}\lambda_2(t) = \max\{q_2(t) - q_3(t) - C_2, -\frac{\lambda_2(t)}{\epsilon}\}, \tag{5.2b}$$

where $\epsilon = \lim_{\Delta \to 0^+} \Delta t$ is an infinitesimal number and equals Δt in the discrete form. To be more specific, for the HOT lane, the demand is $f_1(t) = q_1(t) + q_3(t)$, and the supply is a discontinuous function:

$$g_1(\lambda_1(t)) = \begin{cases} q_1(t) + q_3(t) & \lambda_1(t + \epsilon) = 0, \\ \\ C_1 & \lambda_1(t + \epsilon) > 0. \end{cases} \tag{5.3}$$

For the GP lane, the demand is $f_2(t) = q_2(t) - q_3(t)$, and the supply is

$$g_2(\lambda_2(t)) = \begin{cases} q_2(t) - q_3(t) & \lambda_2(t + \epsilon) = 0, \\ \\ C_2 & \lambda_2(t + \epsilon) > 0. \end{cases} \tag{5.4}$$

Then, the queuing times on HOT and GP lane are:

$$w_1(t) = \frac{\lambda_1(t)}{C_1}, \tag{5.5a}$$

$$w_2(t) = \frac{\lambda_2(t)}{C_2}. \tag{5.5b}$$

Then, after obtaining the queuing times on two types of lanes, the travel time difference, denoted by $w(t)$, can be calculated:

$$w(t) = w_2(t) - w_1(t) = \frac{\lambda_2(t)}{C_2} - \frac{\lambda_1(t)}{C_1}. \tag{5.6}$$

5.2.3 Lane choice model

Intuitively, the proportion of SOVs choosing the HOT lanes depends on the queue length on the two types of lanes, the price and the SOVs' social and economic characteristics. Here we represent the lane choice model by the following functional relationship:

$$q_3(t) = G(\lambda_1(t), \lambda_2(t), u(t), q_2(t)).$$

Reversely, the dynamic prices can be written as:

$$u(t) = G^{-1}(\lambda_1(t), \lambda_2(t), q_3(t), q_2(t)).$$

In the following section, we discuss two examples of lane choice models.

A logit model

If SOVs choose the HOT lanes based on the logit model, then

$$Pr(t) = \frac{1}{1 + \exp(u(t) - \pi^*(\frac{\lambda_2(t)}{C_2} - \frac{\lambda_1(t)}{C_1}))}. \tag{5.7}$$

Here the cost on the HOT and GP lanes are respectively $u(t) + \pi^* \frac{\lambda_1(t)}{C_1}$ and $\pi^* \frac{\lambda_2(t)}{C_2}$. In addition, π^* is the average VOT of all SOVs.

Correspondingly, the price can be written as

$$u(t) = \pi^*(\frac{\lambda_2(t)}{C_2} - \frac{\lambda_1(t)}{C_1}) + \ln\frac{q_2(t) - q_3(t)}{q_3(t)}. \tag{5.8}$$

The vehicle-based user equilibrium principle

In the vehicle-based UE principle (Jin et al., 2018), if a single SOV chooses a lane (HOT or GP), then the cost on the lane is less than or equal to that on the other lane (GP or HOT) which would be experienced by the same vehicle. The vehicle-based UE principle can be easily applied with explicit VOTs and congestion pricing.

At t, the proportion of SOVs choosing the HOT lanes is given by

$$Pr(t) = 1 - F\left(\frac{u(t)}{\frac{\lambda_2(t)}{C_2} - \frac{\lambda_1(t)}{C_1}}\right), \tag{5.9}$$

where $F(\cdot)$ is the cumulative distribution function of $f(\cdot)$.

If the SOVs' VOTs follow a negative exponential distribution, $F(x) = 1 - e^{-\frac{x}{\pi^*}}$, where π^* is the average value of time, then the price can be set as:

$$u(t) = \pi^*\left(\frac{\lambda_2(t)}{C_2} - \frac{\lambda_1(t)}{C_1}\right)\ln\frac{q_2(t)}{q_3(t)}. \tag{5.10}$$

5.2.4   The complete model

Furthermore, the price cannot be negative. Then, combining the objective and constraints above, the full model can be formulated as follows:

$$\text{max} \qquad R = \int_0^T q_3(t)u(t)dt$$

subject to

$$\frac{d}{dt}\lambda_1(t) = \max\{q_1(t) + q_3(t) - C_1, -\frac{\lambda_1(t)}{\epsilon}\}, \tag{5.11}$$
$$\frac{d}{dt}\lambda_2(t) = \max\{q_2(t) - q_3(t) - C_2, -\frac{\lambda_2(t)}{\epsilon}\},$$
$$q_3(t) = q_2(t)G(\lambda_1(t), \lambda_2(t), u(t)),$$
$$u(t) \geq 0.$$

5.3   Approximate analytic solutions with constant demand

Before presenting a method to determine the dynamic price for time-dependent arrival rates, we consider the simpler case when the arrival rates are time-independent. We assume that the HOV flow-rate is constant at $q_1 < C_1$, and the SOV flow-rate is constant at $q_2 > C_2 + C_1 - q_1$. Therefore the total arrival rates $q_1 + q_2$ is greater than the total capacity $C_1 + C_2$. So, the price changes with time t. We assume that the optimal demand of paying SOVs is constant, then $q_3(t) = q_3$. Then, we can formulate an optimization problem.

5.3.1   An optimization problem with the logit model

When a logit model is applied, (5.11) is rewritten as:

$$\max_{q_3} \quad R = \int_0^T q_3 \left(\left(\frac{\lambda_2(t)}{C_2} - \frac{\lambda_1(t)}{C_1} \right) \pi^* + \ln \frac{q_2 - q_3}{q_3} \right) dt$$

subject to

$$\frac{d}{dt} \lambda_1(t) = \max\{q_1 + q_3 - C_1, -\frac{\lambda_1(t)}{\epsilon}\},$$

$$\frac{d}{dt} \lambda_2(t) = \max\{q_2 - q_3 - C_2, -\frac{\lambda_2(t)}{\epsilon}\},$$

$$u(t) \geq 0.$$

$$(5.12)$$

Then, let's consider two cases based on the queue length on the HOT lanes.

- $q_1 + q_3 > C_1$:

  In order to gain revenue, the queuing time on the HOT lanes has to be less than that on the GP lanes. Mathematically, q_3 needs to satisfy:

$$q_3 < \frac{q_2 C_1 - C_2 q_1}{C_1 + C_2}. \tag{5.13}$$

  The dynamic price at time t is $u(t) = \left(\frac{q_2 - q_3 - C_2}{C_2} - \frac{q_1 + q_3 - C_1}{C_1}\right)\pi^* t + \ln\left(\frac{q_2 - q_3}{q_3}\right)$. The objective function is re-written as:

$$\max_{q_3} R = \int_0^T q_3 \left(\frac{q_2 - q_3 - C_2}{C_2} - \frac{q_1 + q_3 - C_1}{C_1} \right) \pi^* t + \ln\left(\frac{q_2 - q_3}{q_3}\right) dt. \tag{5.14}$$

  The first part inside the integral part of (5.14) increases with time, so it will dominate

the function at a very large time t. Then, we want to find a q_3 that maximizes:

$$\Phi_1(q_3) = q_3\left(\frac{q_2 - q_3 - C_2}{C_2} - \frac{q_1 + q_3 - C_1}{C_1}\right)$$

The optimal q_3, denoted as q_3^*, is:

$$q_3^* = \frac{C_1 q_2 - C_2 q_1}{2(C_1 + C_2)} \tag{5.15}$$

- $q_1 + q_3 \leq C_1$.

   There will be no queue on the HOT lane, $u(t) = \frac{q_2 - q_3 - C_2}{C_2}\pi^* t + \ln(\frac{q_2 - q_3}{q_3})$. Then, the objective function becomes:

$$\max_{q_3} R = \int_0^T q_3 \frac{q_2 - q_3 - C_2}{C_2}\pi^* t + \ln(\frac{q_2 - q_3}{q_3}) dt.$$

   Similar to the analysis above, the optimal q_3 maximizes the following function:

$$\Phi_2(q_3) = q_3 \frac{q_2 - q_3 - C_2}{C_2}$$

   Then, q_3^* is calculated by

$$q_3^* = \frac{q_2 - C_2}{2} \tag{5.16}$$

Then, we need to check the relationship between the demand pattern and the maximum revenue. We evaluate the function $\Phi(q_3)$ at three critical points for q_3: $C_1 - q_1$, $\frac{q_2 - C_2}{2}$, and $\frac{C_1 q_2 - C_2 q_1}{2(C_1 + C_2)}$.

1. $\frac{q_2 - C_2}{2} \leq C_1 - q_1$.

   In this case, $q_2 + 2q_1 \leq 2C_1 + C_2$, and $\frac{C_1 q_2 - C_2 q_1}{2(C_1 + C_2)}$ is automatically less than $C_1 - q_1$. We

93

just need to evaluate $\Phi_2(q_3)$ at $q_3 = \frac{q_2 - C_2}{2}$ and $q_3 = C_1 - q_1$ to get the maximum value.

2. $\frac{C_1 q_2 - C_2 q_1}{2(C_1 + C_2)} > C_1 - q_1$:

In this case, $q_2 + (2 + \frac{C_2}{C_1})q_1 > 2(C_1 + C_2)$, thus $\frac{q_2 - C_2}{2} > C_1 - q_1$. We just need to compare $\Phi_1(q_3)$ at $q_3 = \frac{C_1 q_2 - C_2 q_1}{2(C_1 + C_2)}$ and $q_3 = C_1 - q_1$ to get the maximum value.

5.3.2 An optimization problem with the UE principle

When the UE principle is applied, (5.11) is rewritten as:

$$\max_{q_3} \qquad R = \int_0^T q_3 \pi^* (\frac{\lambda_2(t)}{C_2} - \frac{\lambda_1(t)}{C_1}) \ln \frac{q_2}{q_3} dt$$

subject to

$$\frac{d}{dt}\lambda_1(t) = \max\{q_1 + q_3 - C_1, -\frac{\lambda_1(t)}{\epsilon}\}, \tag{5.17}$$

$$\frac{d}{dt}\lambda_2(t) = \max\{q_2 - q_3 - C_2, -\frac{\lambda_2(t)}{\epsilon}\},$$

$$u(t) \geq 0.$$

Same as previous analysis, we consider two scenarios.

- $q_1 + q_3 > C_1$.

  In addition, q_3 should satisfy the constraint: $q_3 < \frac{q_2 C_1 - C_2 q_1}{C_1 + C_2}$. The dynamic price at time t is $u(t) = (\frac{q_2 - q_3 - C_2}{C_2} - \frac{q_1 + q_3 - C_1}{C_1})\pi^* t \ln(\frac{q_2}{q_3})$. The objective function is re-written as:

  $$\max_{q_3} R = \int_0^T q_3 (\frac{q_2 - q_3 - C_2}{C_2} - \frac{q_1 + q_3 - C_1}{C_1})\pi^* t \ln(\frac{q_2}{q_3}) dt.$$

Equivalently, we want to maximize the following function:

$$\Phi_3(q_3) = q_3 \big(\frac{q_2 - q_3 - C_2}{C_2} - \frac{q_1 + q_3 - C_1}{C_1} \big) \ln \big(\frac{q_2}{q_3} \big).$$

q_3^* is reached when $\frac{\mathrm{d}\Phi_3(q_3^*)}{\mathrm{d}q_3^*} = 0$, i.e.,

$$\big(\frac{q_2 - C_2 - 2q_3^*}{C_2} - \frac{q_1 + 2q_3^* - C_1}{C_1} \big) \ln \frac{q_2}{q_3^*} - \big(\frac{q_2 - q_3^* - C_2}{C_2} - \frac{q_1 + q_3^* - C_1}{C_1} \big) = 0 \quad (5.18)$$

- $q_1 + q_3 \leq C_1$.

  There will be no queue on the HOT lane, $u(t) = \frac{q_2 - q_3 - C_2}{C_2} \pi^* t \ln(\frac{q_2}{q_3})$. Then, the objective function becomes:

$$\max_{q_3} R = \int_0^T q_3 \frac{q_2 - q_3 - C_2}{C_2} \pi^* t \ln \big(\frac{q_2}{q_3} \big) dt.$$

  Equivalently, we want to maximize the following function:

$$\Phi_4(q_3) = q_3 \frac{q_2 - q_3 - C_2}{C_2} \ln \big(\frac{q_2}{q_3} \big). \tag{5.19}$$

  Then q_3^* is reached when $\frac{\mathrm{d}\Phi_4(q_3^*)}{\mathrm{d}q_3^*} = 0$, i.e.,

$$(q_2 - C_2 - 2q_3^*) \ln \frac{q_2}{q_3^*} - (q_2 - q_3^* - C_2) = 0 \tag{5.20}$$

Similar to Section 5.3.1, we need to evaluate $\Phi_3(q_3)$ and $\Phi_4(q_3)$ at those three critical points.

95

5.4 Solutions with an optimal control approach

5.4.1 Formulation of the optimal control problem

We formulate two optimal control problem with mixed inequality constraints based on the switching system (5.2a).

Let $\mu_1(t)$ and $\mu_2(t)$ be the co-state variables associated with the changing rate of the queue lengths on the HOT and GP lanes. Then, we define the Hamiltonian function as:

$$H(\lambda_1, \lambda_2, \mathbf{q_3}, t) = q_3(t)G^{-1}(\lambda_1(t), \lambda_2(t), q_3(t)) + \mu_1(t)(f_1(t) - g_1(\lambda_1(t))) + \mu_2(t)(f_2(t) - g_2(\lambda_2(t))).$$

$$(5.21)$$

According to the Pontryagin minimum principle (PMP), when the minimum of the above optimal control problem is reached, the queue lengths on both lanes should be:

$$\frac{d}{dt}\lambda_1^*(t) = \max\{q_1(t) + q_3^* - C_1, -\frac{\lambda_1^*(t)}{\epsilon}\}, \tag{5.22a}$$

$$\frac{d}{dt}\lambda_2^*(t) = \max\{q_2(t) - q_3^* - C_2, -\frac{\lambda_2^*(t)}{\epsilon}\}. \quad t \in [0, T]. \tag{5.22b}$$

The initial conditions are

$$\lambda_1^*(0) = \lambda_1(0), \tag{5.23a}$$

$$\lambda_2^*(0) = \lambda_2(0). \tag{5.23b}$$

For the co-state equations:

$$\frac{d}{dt}\mu_1(t) = -\frac{\partial H}{\partial \lambda_1^*(t)},$$ (5.24a)

$$\frac{d}{dt}\mu_2(t) = -\frac{\partial H}{\partial \lambda_2^*(t)}. \quad t \in [0, T].$$ (5.24b)

This optimal problem has a fixed final time but free final states, it is straightforward that

$$\mu_1(T) = 0,$$ (5.25a)

$$\mu_2(T) = 0.$$ (5.25b)

For the optimal dynamic prices, $\frac{\partial H}{\partial q_3^*(t)}$ is required, which is equivalent to

$$\frac{\partial H(\lambda_1^*, \lambda_2^*, \mathbf{q_3^*}, t)}{\partial q_3^*(t)} = 0$$ (5.26)

With (5.22), 5.24, (5.23), and (5.25), we can formulate a two-point boundary-value problem (TPBVP). Since $g_1(\lambda_1(t))$ and $g_2(\lambda_2(t))$ are discontinuous functions, accurate solutions cannot be achieved by continuous BVP solvers, such as bvp4c command in the MATLAB (Nikoobin and Moradi, 2017).

5.4.2   An iterative method

Many iterative methods have been considered for the solution of the TPBVP problem. A fixed step size is chosen for computational simplicity in (Yang and Huang, 1997). The steepest descent method, variation of extremals and quasilinearization were considered in (Kirk, 2012).

In this section, we provide an iterative method to solve the optimal control problem. The algorithm is described as follows:

- Step 1: Choose the initial values of the control variables $q_3^n(t)$. Let the iteration index i be zero.

- Step 2: With the nominal control history $\mathbf{q_3}(t)$, integrate (5.2) from 0 to T with initial conditions ($\lambda_1(0) = 0$ and $\lambda_2(0) = 0$).

- Step 3: Integrate the co-state equations from T to 0, with "initial" conditions ($\mu_1(T) = 0$ and $\mu_2(T) = 0$).

- Step 4: Evaluate the gradient $\frac{\partial H^{(i)}}{\partial \mathbf{q_3}(t)}$.

- Step 5: Update the prices by

$$\mathbf{q_3}^{i+1}(t) = \mathbf{q_3}^{(i)}(t) - \alpha \frac{\partial H^{(i)}}{\partial \mathbf{q_3}(t)}.$$

  where α is a fixed step size. If the change in the Hamilton between successive iterations is sufficiently small, stop; otherwise replace $\mathbf{q_3}^{(i)}(t)$ by $\mathbf{q_3}^{(i+1)}(t)$, $k = 0, 1, ..., N - 1$, and return to step 2.

5.5   Numeric results

For our numerical analysis, the study site is a freeway segment with a lane-drop downstream of the GP lane. The capacity for a single HOT and a single GP lane is the same, with the value of 30 veh/min ($C_1 = C_2 = 30$ veh/min). An one-hour period is studied, and the time interval is one minute. We set the initial queues to zero ($\lambda_1(0) = 0$, $\lambda_2(0) = 0$). The average VOT is \$0.5/min. We test two scenarios: (1) the demand of HOVs is constant at $q_1(t) = 10$

veh/min, and the demand of SOVs is constant at $q_2(t) = 60$ veh/min; and (2) the demand of HOVs is constant at $q_1(t) = 25$ veh/min, and the demand of SOVs is constant at $q_2(t) = 60$ veh/min. Our initial guess of $q_3(t)$ is 10 veh/min.

5.5.1 The logit model

When the lane group choice is capture by a logit model, the dynamic price is calculated by (5.8), and the revenue during the study period is

$$R = \int_0^T q_3(t)((\frac{\lambda_2(t)}{C_2} - \frac{\lambda_1(t)}{C_1})\pi^* + \ln(\frac{q_2(t) - q3(t)}{q_3(t)}))dt. \tag{5.27}$$

When the demand of HOVs is 10 veh/min, the system will reach the states as shown in Figure 5.1. There is no queue on the HOT lanes, and the price linearly increase with time. The total revenue for the one-hour study period is \$7628.

(a) Performance (b) Revenue

Figure 5.1: Numerical results with the logit model (low demand of HOVs).

When the demand of HOVs is high, queue forms on the HOT lanes when the maximum revenue is gained. The operators can gain revenue of \$5450.

(a) Performance   (b) Revenue

Figure 5.2: Numerical results with the logit model (high demand of HOVs).

5.5.2 The vehicle-based UE principle

When the lane group choice is capture by the vehicle-based UE principle with negative exponentially distributed VOTs, the dynamic price is calculated by (5.10), and the revenue during the study period is

$$R = \int_0^T q_3(t)\left(\frac{\lambda_2(t)}{C_2} - \frac{\lambda_1(t)}{C_1}\right)\pi^* \ln\left(\frac{q_2(t)}{q_3(t)}\right)dt. \tag{5.28}$$

When the demand of HOVs is 10 veh/min, there is no queue on the HOT lanes, as shown in Figure 5.3. The total revenue for the one-hour study period is \$10588.

(a) Performance   (b) Revenue

Figure 5.3: Numerical results with the UE principle (low demand of HOVs).

When the demand of HOVs is high, the queue lengths on both lanes keep increasing during the study period. The revenue is $9382 after operating the HOT lanes for one hour.

(a) Performance   (b) Revenue

Figure 5.4: Numerical results with the UE principle (high demand of HOVs).

5.6 Comparison with different objectives

In this section, we evaluate the "performance-optimization" pricing schemes (in Chapter 4) with three metrics: the queue lengths on the HOT lanes, the throughput of the freeway, and the revenue for the operators. When the freeway performance is optimized, no queue exists on the HOT lanes, and the throughput of the HOT lanes is at its capacity. The two demand patterns are the same as Section 5.5.

5.6.1 The logit model

When the lane group choice is capture by a logit model, the dynamic price is calculated by (5.8), and the revenue during the study period is

$$R = \int_0^T q_3(t)(\frac{q_2(t) - q_3(t) - C_2}{C_2}\pi^* t + \ln(\frac{q_2(t) - q3(t)}{q_3(t)}))dt. \tag{5.29}$$

When $q_1(t) = 10$ veh/min, based on the optimal state defined in (3.12) and (3.13), the optimal demand of paying SOVs is $q_3^*(t) = 20$ veh/min. The queue length on the HOT lane, $\lambda_1(t)$ is 0, and the throughputs on the HOT and GP lanes are both 30 veh/min, which is the capacity of one lane. Based on (5.29), the total revenue during the one-hour period is \$6731.

The optimal demand of paying SOVs is $q_3^*(t) = 5$ veh/min when the demand of HOVs is high. The queue length on the HOT lane, $\lambda_1(t)$ is 0, and the throughputs on the HOT and GP lanes are both 30 veh/min. The total revenue during the one-hour period is \$4407.

5.6.2 The vehicle-based UE principle

When the lane group choice is capture by the vehicle-based UE principle with negative exponentially distributed VOTs, the dynamic price is calculated by (5.10), and the revenue during the study period is

$$R = \int_0^T q_3(t) \frac{q_2(t) - q_3(t) - C_2}{C_2} \pi^* t \ln(\frac{q_2(t)}{q_3(t)}) dt. \tag{5.30}$$

The optimal demand of paying SOVs is $q_3^*(t) = 20$ veh/min with low demand of SOVs. The queue length on the HOT lane, $\lambda_1(t)$ is 0, and the throughputs on the HOT and GP lanes are both 30 veh/min, which is the capacity of one lane. From (5.30), the total revenue during the one-hour period is \$6482.

The optimal demand of paying SOVs is $q_3^*(t) = 5$ veh/min when the HOVs' demand is high. The queue length on the HOT lane, $\lambda_1(t)$ is 0, and the throughputs on the HOT and GP lanes are both 30 veh/min. The total revenue during the one-hour period is \$9163.

5.6.3 A comparison of pricing schemes

In this section, we numerically compare the the freeway performance and the revenue for the operators under different pricing schemes. When the demand of HOVs is low, the corresponding metrics are summarized in Table 5.1.

Table 5.1: Comparison of two pricing schemes with low demand of HOVs

	Performance-optimization pricing schemes		Revenue maximization pricing schemes	
Lane choice model	Logit	UE	Logit	UE
$\lambda_1(T)$	0	0	0	0
$g_1(t)$	30	30	25	19
Revenue	6731	6481	7628	10588

From Table 5.1, we find that in order to maximize the revenue, the operators need to set a higher price to discourage SOVs from switching to the HOT lanes. In this case, the HOT lanes are utilized, and the vehicles on the HOT lanes can move at free-flow speed. We also check the scenario with high demand of HOVs.

Table 5.2: Comparison of two pricing schemes with high demand of HOVs

	Performance-optimization pricing schemes		Revenue maximization pricing schemes	
Lane choice model	Logit	UE	Logit	UE
$\lambda_1(T)$	0	0	237	75
$g_1(t)$	30	30	30	30
Revenue	4407	9163	5450	9382

From 5.2, since more SOVs are willing to pay the tolls, the HOT lanes are congested when the maximum revenue is obtained.

Based on the numerical results with both low and high demand patterns, we conclude that revenue maximization may not be consistent with the performance optimization for the HOT lane operations. This is not the same as what is stated in (Perez and Sciara, 2003).

5.7 Conclusion

In this study, we try to find the optimal price that maximizes the revenue for the HOT lanes when the freeway is congested while the demand of the HOVs is lower than the capacity of the HOT lanes. We apply a point queue model to capture the traffic dynamics. We model the lane choice behavior of SOVs with a logit model and a vehicle-based user equilibrium principle. The traffic model and the lane choice model are the constraints for the problem. We first consider a constant demand case, and find an approximate solution for the optimization problem. Then, we formulate an optimal control problem, and propose an iterative algorithm to find the optimal dynamic price. Numerical results are provided for different lane choice models and different demand patterns. Then, we compare them with the numerical results for the case when the freeway performance is optimized. We choose queue length and throughput on the HOT lanes, and the revenue for the HOT lanes as metrics. We find that when the traffic demand is not too high, we can generate maximum revenue when the HOT lane is at free-flow condition; when the demand of HOVs is close to capacity of the HOT lane, we can get maximum revenue when the total travel time is minimized. So, the revenue maximization may not coincide with the performance optimization for the HOT lanes.

The following are some potential future research topics.

- In this study, we assume the demand of HOVs is lower than the capacity and the total demand is higher than the freeway's capacity. In the future study, we want to maximize the in a different traffic condition: $\int_0^T q_1(s) + q_2(s)ds < (C_1 + C_2)T$ and $q_1(t) < C_1$.

- After realizing that performance optimization and revenue maximization can't be achieved at the same time, we are interested in design pricing schemes that achieve some mixed objectives. For example, what would the pricing scheme be when the revenue for the HOT lanes is maximized the speed is over 45 mph on the HOT lanes.

105

- In this study, we assume the operators have full information of the lane choice behavior of SOVs. We want to investigate if it is possible to learn SOVs' lane choice model while maximizing the revenue for the HOT lanes.

Chapter 6

Pricing Schemes for High-occupancy Toll Lanes Considering the Departure Time User Equilibrium

6.1   Introduction

In Chapter 3 and 4, there is one assumption that SOVs only make choice about the lane group when approaching the entrance of the HOT lanes. In those cases, only the lane group choice is considered in the behavior modeling. However, in real world, travelers also choose their departure time, especially during the morning and evening peak periods.

When the departure time equilibrium is reached, no vehicle wants to depart further early or late, and the total cost for each vehicle on either HOT or GP lanes is the same (but the cost for all vehicles can be different). For each vehicle, the total cost is balanced by a high scheduling cost and a low queuing time cost, or vice versa. In the presence of HOT lanes, SOVs can balance their total cost by adjusting the departure time as well as the lane group

choice.

In most HOT lane studies, a common assumption is that the demand of HOVs is lower than the capacity of the HOT lanes. A full-utilization principle is applied to optimize the freeway performance. When considering the departure time equilibrium, the HOT lanes are already congested during the peak period, thus the above full-utilization principle is not applicable. So, we are interested in exploring what pricing schemes would be applicable to the HOT lanes. The rest of this study is organized as follows. In Section 6.2, we describe the dynamics of the traffic system and define the cost functions for different types of vehicles. In Section 6.3, we show the departure time user equilibrium analysis with and without tolls. In Section 6.4, we propose three flat pricing schemes that meet the following criteria: (1) the total travel time and scheduling cost is minimized; and (2) the costs for each non-paying and paying SOV are the same. The difference between those schemes exists is caused by the additional criteria: the revenue collected by operators and the tolling constrains for certain type of vehicles. In Section 6.5, we present a case study to show the departure time equilibrium and the performance with five metrics in different pricing schemes. In Section 6.6, we compare pricing schemes with and without consideration of departure time choice. In Section 6.7, we conclude this study with potential extensions.

6.2 Problem Formulation

The study site is the same as Chapter 3. But different from Chapter 3 and 4, SOVs can choose both the lane group and the departure time when the pricing scheme is available. In this system, HOVs and the paying SOVs share the HOT lanes, but non-paying SOVs stay on the GP lanes. During the congested period, there are two parallel bottlenecks, one for each lane group.

6.2.1 Definitions of variables and the point queue model

Figure 6.1 is an illustration of point queue models for a traffic system with the HOT and GP lanes. We define the road and traffic flow characteristics with the following variables:

- Traffic demands: N_1 and N_2 are the travel demands of HOVs and SOVs during the peak period.

- Capacities: C_1 and C_2 are the bottleneck capacities of the HOT and GP lanes.

- Departure cumulative flows: $F_1(t)$ and $F_2(t)$ are respectively the departure cumulative flows on the HOT and GP lanes at time t.

- Departure flow-rates: $f_1(t)$ and $f_2(t)$ are the departure flow-rates

- Queue lengths: $\lambda_1(t)$ and $\lambda_2(t)$ are the queue lengths on the HOT and GP lanes, respectively.

- Queuing times: $w_1(t)$ and $w_2(t)$ are respectively the queuing times for vehicles departing at t on the HOT and GP lanes.

When pricing schemes are available, we denote the demand of paying SOVs by N_3. Thus, the demands on the HOT and GP lanes are respectively $N_1 + N_3$ and $N_2 - N_3$. Meanwhile, the average occupancy of HOVs is σ_1.

Figure 6.1: Point queue representation of the traffic system with the HOT lanes and definitions of traffic flow variables.

6.2.2 Costs for vehicles

For simplicity, we assume the free-flow travel time is zero. Then, the cost for a vehicle at a single bottleneck comprises of three parts: the queuing cost, caused by congestion, the scheduling cost, caused by the schedule delay, and the toll.

First, the queuing cost for a vehicle departing at time t is:

$$\Upsilon_1(t) = \alpha w(t).$$

where α represents the VOT. For the scheduling cost, we assume all the drivers have the same ideal arrival time, t^*. A piece-wise linear scheduling cost function for a driver departing at time t can be written as:

$$\Upsilon_2(t) = \beta \max\{t^* - t - w(t), 0\} + \gamma \max\{t - t^* + w(t), 0\},$$

where β and γ represent the penalty for an early arrival and a late arrival, respectively. It is usually assumed that $\beta < \alpha$, in order to guarantee the existence of user equilibrium. In general, $\beta/\alpha = 0.5$, and $\gamma/\alpha = 2$ (Small, 2015). We denote the toll for a vehicle by $u(t)$, the vehicle's total cost can be described by the following model:

$$\phi(t) = \alpha w(t) + \beta \max\{t^* - t - w(t), 0\} + \gamma \max\{t - t^* + w(t), 0\} + u(t).$$

In this study, we assume α, β, and γ are constant. Furthermore, the tolls for each HOV, non-paying SOV, and paying SOV are respectively $u_1(t)$, $u_2(t)$, and $u_3(t)$ when a pricing scheme is applied. Then, the costs for each HOV, non-paying SOV and paying SOV departing at time t are respectively:

$$\phi_1(t) = \alpha w_1(t) + \beta \max\{t^* - t - w_1(t), 0\} + \gamma \max\{t - t^* + w_1(t), 0\} + u_1(t), \quad \text{(6.1a)}$$

$$\phi_2(t) = \alpha w_2(t) + \beta \max\{t^* - t - w_2(t), 0\} + \gamma \max\{t - t^* + w_2(t), 0\} + u_2(t), \quad \text{(6.1b)}$$

$$\phi_3(t) = \alpha w_1(t) + \beta \max\{t^* - t - w_1(t), 0\} + \gamma \max\{t - t^* + w_1(t), 0\} + u_3(t). \quad \text{(6.1c)}$$

6.3 Departure time user equilibrium

6.3.1 Without tolls

In this section, we analyze a case when pricing schemes are not applied to the HOT lanes. In this scenario, the HOT lanes are equivalent to the HOV lanes, in which SOVs are only allowed to use the GP lanes. Thus $N_3 = 0$, and the demands on the HOT and GP lanes are N_1 and N_2, respectively. We denote t_i^e and t_i^s $(i = 1, 2)$ as the earliest and latest departure times on the HOT and GP lanes, respectively. Here, we define the "underutilized" HOT lanes as follows: the HOT lanes are underutilized when the congestion period on the HOT lanes is shorter than that for the GP lanes. The bold lines on the x-axis in Figure 6.2 show the periods when the HOT lanes are underutilized.

We further denote $\tilde{t}_i$ $(i = 1, 2)$ as the departure time for vehicles arriving at t^* on the HOT and GP lanes, respectively. We let $f_{11}(t)$ and $f_{12}(t)$ be the departure flow-rates for $t \in [t_1^s, \tilde{t}_1]$ and $t \in (\tilde{t}_1, t_1^e]$ for the HOT lane, respectively; $f_{21}(t)$ and $f_{22}(t)$ be the departure flow-rates for $t \in [t_2^s, \tilde{t}_2]$ and $t \in (\tilde{t}_2, t_2^e]$ for the GP lanes, respectively. Since both lane groups are fully utilized during the congested period, the corresponding arrival flow-rates are $g_1(t) = C_1$ and $g_2(t) = C_2$ based on the point queue model.

Figure 6.2: The equilibrium without tolls

The equilibrium state analysis is similar to (Arnott et al., 1990a). For the HOT lanes, the costs for vehicles departing at t_1^s, $\tilde{t}_1$ and t_1^e are:

$$\phi_{t_1^s} = \beta(t^* - t_1^s), \tag{6.2a}$$

$$\phi_{\tilde{t}_1} = \alpha w_1(\tilde{t}_1) + \beta(t^* - \tilde{t}_1 - w_1(\tilde{t}_1)), \tag{6.2b}$$

$$\phi_{t_1^e} = \gamma(t_1^e - t^*). \tag{6.2c}$$

For vehicles departing at t_1^s and $\tilde{t}_1$, their costs should be the same at the equilibrium state,

i.e. $\phi_{t_1^s} = \phi_{\tilde{t}_1}$, which leads to:

$$\beta(t^* - t_1^s) = \alpha w_1(\tilde{t}_1) + \beta(t^* - \tilde{t}_1 - w_1(\tilde{t}_1)), \tag{6.3}$$

where $w_1(\tilde{t}_1) = \frac{(f_{11}(t) - C_1)(\tilde{t}_1 - t_1^s)}{C_1}$. We can derive $f_{11}(t)$ based on (6.3):

$$f_{11}(t) = C_1 + \frac{C_1 \beta}{\alpha - \beta} = \frac{\alpha}{\alpha - \beta} C_1 \qquad t \in [t_1^s, \tilde{t}_1]. \tag{6.4}$$

At the same time, vehicles departing at t^* and t_1^e should have the same cost, i.e., $\phi_{t^*} = \phi_{t_1^e}$. Then, we have:

$$\alpha w_1(t^*) + \gamma w_1(t^*) = \gamma(t_1^e - t^*), \tag{6.5}$$

where $w_1(t^*) = \frac{(C_1 - f_{12}(t))(t_1^e - t^*)}{C_1}$. We can get $f_{12}(t)$ from $w_1(t^*)$ and (6.5):

$$f_{12}(t) = C_1 - \frac{C_1 \gamma}{\alpha + \gamma} = \frac{\alpha}{\alpha + \gamma} C_1 \qquad t \in (\tilde{t}_1, t_1^e]. \tag{6.6}$$

Meanwhile, the cost for the first and the last vehicle departing during the congested period should be identical as well, which means $\beta(t^* - t_1^s) = \gamma(t_1^e - t^*)$. From Figure 6.2, it is straightforward that the length of the congestion period, or equivalently, the time needed to clear the queue, equals $t_1^e - t_1^s = \frac{N_1}{C_1}$, we can further calculate the starting and ending times for the congested period (i.e., the earliest and latest departure times in the equilibrium state):

$$t_1^s = t^* - \frac{\gamma}{\beta + \gamma} \frac{N_1}{C_1}, \tag{6.7a}$$

$$t_1^e = t^* + \frac{\beta}{\beta + \gamma} \frac{N_1}{C_1}. \tag{6.7b}$$

From (6.7), we find that the starting and ending times for the congested period are inde-

114

pendent of the VOT. We can also get $\tilde{t}_1$ based on the condition, $\phi_1(t_1^s) = \phi_1(\tilde{t}_1)$:

$$\tilde{t}_1 = t^* - \frac{\beta}{\alpha} \frac{\gamma}{\beta + \gamma} \frac{N_1}{C_1}. \tag{6.8}$$

So, when the system reaches departure time equilibrium, a fraction of $\frac{\gamma}{\beta+\gamma}$ of HOVs departs early, and the rest departs late. Meanwhile, the cost for each HOV is independent of the VOT, as shown in (6.9):

$$\phi_1(t) = \frac{\beta\gamma}{\beta + \gamma} \frac{N_1}{C_1}. \tag{6.9}$$

Similarly, during the peak period, the departure flow-rates, critical times and cost for each SOV on the GP lanes can be calculated by (6.10):

$$f_{21}(t) = \frac{\alpha}{\alpha - \beta} C_2 \qquad t \in [t_2^s, \tilde{t}_2], \tag{6.10a}$$

$$f_{22}(t) = \frac{\alpha}{\alpha + \gamma} C_2 \qquad t \in (\tilde{t}_2, t_2^e], \tag{6.10b}$$

$$t_2^s = t^* - \frac{\gamma}{\beta + \gamma} \frac{N_2}{C_2}, \tag{6.10c}$$

$$t_2^e = t^* + \frac{\beta}{\beta + \gamma} \frac{N_2}{C_2}, \tag{6.10d}$$

$$\tilde{t}_2 = t^* - \frac{\beta}{\alpha} \frac{\gamma}{\beta + \gamma} \frac{N_2}{C_2}, \tag{6.10e}$$

$$\phi_2(t) = \frac{\beta\gamma}{\beta + \gamma} \frac{N_2}{C_2}. \tag{6.10f}$$

When we assume identical α, β, and γ for all drivers, the choice of earliest and latest departure times only depends on the total demand and the capacity of each lane group. When comparing the two costs in (6.9) and (6.10f), we find that there is no incentive for SOVs to pay to choose the HOT lanes if $\frac{N_1}{C_1} > \frac{N_2}{C_2}$. In the following section, we analyze the

equilibrium state with the following demand profile: $\frac{N_1}{C_1} \leq \frac{N_2}{C_2}$.

6.3.2 With tolls

When $\frac{N_1}{C_1} \leq \frac{N_2}{C_2}$, the duration of the congested period for the GP lanes is longer than the HOT lanes. In order to arrive closer to the desired time, some SOVs are willing to pay and switch to the HOT lanes. When N_3 SOVs switch to the HOT lane, the cost for the paying and non-paying SOVs should be the same based on the user equilibrium principle.

Figure 6.3: The equilibrium with tolls

From Figure 6.3, we can see that "underutilized" period for the HOT lanes is shorter than that in Figure 6.2. We denote τ_1^s, τ_1^e, and $\tilde{\tau}_1$ as the earliest and latest departure times, and the departure time for arriving at the desired time for the HOT lanes, respectively. Similar to the earlier analysis, we can derive three critical times for the HOT lanes:

$$\tau_1^s = t^* - \frac{\gamma}{\beta + \gamma} \frac{N_1 + N_3}{C_1}, \tag{6.11a}$$

$$\tau_1^e = t^* + \frac{\beta}{\beta + \gamma} \frac{N_1 + N_3}{C_1}, \tag{6.11b}$$

$$\tilde{\tau}_1 = t^* - \frac{\beta}{\alpha} \frac{\gamma}{\beta + \gamma} \frac{N_1 + N_3}{C_1}. \tag{6.11c}$$

For the GP lanes in the new equilibrium, the corresponding times are in (6.12):

$$\tau_2^s = t^* - \frac{\gamma}{\beta + \gamma} \frac{N_2 - N_3}{C_2}, \tag{6.12a}$$

$$\tau_2^e = t^* + \frac{\beta}{\beta + \gamma} \frac{N_2 - N_3}{C_2}, \tag{6.12b}$$

$$\tilde{\tau}_2 = t^* - \frac{\beta}{\alpha} \frac{\gamma}{\beta + \gamma} \frac{N_2 - N_3}{C_2}. \tag{6.12c}$$

From the analysis in the previous section, we know that the sum of travel and scheduling costs is constant in the equilibrium state. So, $u_1(t)$, $u_2(t)$ and $u_3(t)$ should also be constant to drive the system to the new departure time user equilibrium. Thus we drop the notation (t) in the tolls in the following sections. Then, the total cost for each HOV, non-paying SOV,

and paying SOVs are:

$$\Phi_1(t) = \frac{\beta\gamma}{\beta+\gamma}\frac{N_1+N_3}{C_1} + u_1, \tag{6.13a}$$

$$\Phi_2(t) = \frac{\beta\gamma}{\beta+\gamma}\frac{N_2-N_3}{C_2} + u_2, \tag{6.13b}$$

$$\Phi_3(t) = \frac{\beta\gamma}{\beta+\gamma}\frac{N_1+N_3}{C_1} + u_3. \tag{6.13c}$$

In the new equilibrium state, the total number of commuters on the HOT lanes is $N_1\sigma_1+N_3$, and the number of commuters on the GP lanes is N_2-N_3. Thus the sum of travel and scheduling costs for a HOV/paying SOV is $\frac{\beta\gamma}{\beta+\gamma}\frac{N_1+N_3}{C_1}$, and the corresponding cost for a non-paying SOV is $\frac{\beta\gamma}{\beta+\gamma}\frac{N_2-N_3}{C_2}$. Then, the total travel time and scheduling costs for all commuters equals:

$$T(N_3) = \frac{\beta\gamma}{\beta+\gamma}(\frac{N_1+N_3}{C_1}(N_1\sigma_1+N_3) + \frac{N_2-N_3}{C_2}(N_2-N_3)), \tag{6.14}$$

The above function shares the following properties:

1. $T(N_3) < T(0)$ for $N_3 > 0$. We take the derivative of the function with respect to N_3:

$$T'(N_3) = \frac{\beta\gamma}{\beta+\gamma}(\frac{2N_3+N_1+N_1\sigma_1}{C_1} + \frac{2(N_3-N_2)}{C_2}). \tag{6.15}$$

SOVs will pay to switch to the HOT lanes only when the sum of their travel cost and scheduling cost decreases.

2. The function should be non-increasing when $N_3 = 0$. From (6.15), we can see:

$$T'(0) = \frac{\beta\gamma}{\beta+\gamma}(\frac{N_1+N_1\sigma_1}{C_1} - \frac{2N_2}{C_2}) \leq 0. \tag{6.16}$$

This reveals the condition when pricing schemes can be applied to minimize the total cost: $\frac{\sigma_1+1}{2}\frac{N_1}{C_1} \leq \frac{N_2}{C_2}$. Note that if $\sigma_1 = 1$, it is the same as the case at the end of Section 6.3.1.

6.4 Three pricing schemes

The first criterion for pricing schemes is introduced as follows:

C1. The total travel time and scheduling costs for all commuters, is minimized. That is, we want to find a N_3 such that:

$$\min_{N_3} T = \frac{\beta\gamma}{\beta+\gamma}(\frac{N_1+N_3}{C_1}(N_1\sigma_1+N_3) + \frac{N_2-N_3}{C_2}(N_2-N_3)). \tag{6.17}$$

Intuitively, (6.17) is minimized when the derivative of the right-hand-side with respect to N_3 is 0, which means:

$$\frac{N_1\sigma_1+N_3+N_1+N_3}{2C_1} = \frac{N_2-N_3}{C_2}.$$

Then, the optimal number of paying SOVs, denoted by N_3^*, is:

$$N_3^* = -\frac{C_2}{C_1+C_2}N_1\frac{\sigma_1+1}{2} + \frac{C_1}{C_1+C_2}N_2. \tag{6.18}$$

During the peak period, the optimal demand of paying SOVs in (6.18) can be achieved with appropriate pricing schemes. In general, the paying behavior of SOVs increases the cost for HOVs, which will discourage the choice of carpooling. So, we need to provide a credit scheme for HOVs, and a pricing scheme for SOVs. In this case, $u_1 \leq 0$, $u_2 \geq 0$, and $u_3 \geq 0$.

The second criterion for pricing schemes is the regarding the total cost for each non-paying

and paying SOV:

C2. The costs for each non-paying and paying SOV are the same. That is, u_2 and u_3 should satisfy:

$$\frac{\beta\gamma}{\beta+\gamma}\frac{N_2 - N_3^*}{C_2} + u_2 = \frac{\beta\gamma}{\beta+\gamma}\frac{N_1 + N_3^*}{C_1} + u_3, \tag{6.19}$$

In the following sections, we will bring in more requirements from the operators, and propose three pricing based on the additional criterion.

6.4.1 Pricing scheme 1

One possible constraint for the system is that,

C3-1. Non-paying SOVs do not need to pay to use the GP lanes, and HOVs will not receive any credit, denoted by u_1 for carpooling. That is $u_1 = 0$ and $u_2 = 0$.

In the corresponding pricing scheme, each paying SOV needs to pay:

$$u_3 = \frac{\beta\gamma}{\beta+\gamma}\left(\frac{N_2 - N_3^*}{C_2} - \frac{N_1 + N_3^*}{C_1}\right). \tag{6.20}$$

The total cost for a non-paying SOV and paying SOV is: $\Phi_2(t) = \Phi_3(t) = \frac{\beta\gamma}{\beta+\gamma}\frac{N_2 - N_3^*}{C_2}$. So, the revenue for operators is:

$$R_1 = \frac{\beta\gamma}{\beta+\gamma}\left(\left(\frac{N_2 - N_3^*}{C_2} - \frac{N_1 + N_3^*}{C_1}\right)N_3^*\right). \tag{6.21}$$

6.4.2　Pricing scheme 2

In the second pricing scheme, we charge non-paying SOVs at a very low price, and paying SOVs at a relatively high price. The pricing scheme should satisfy the following criterion:

C3-2. The total revenue collected from SOVs and the credit provided to HOVs are balanced. That is, we want to get a set of (u_1, u_2, u_3) such that:

$$u_1 N_1 = u_2 N_2 + (u_3 - u_2) N_3. \tag{6.22}$$

Together with the first and second criterion, the tolls for each non-paying and paying SOV are set as:

$$u_2 = \frac{\beta\gamma}{\beta+\gamma} \frac{N_1^{\frac{3-\sigma_1}{2}}(-C_2 N_1^{\frac{\sigma_1+1}{2}} + C_1 N_2)}{C_1 N_2 (C_1 + C_2)}, \tag{6.23a}$$

$$u_3 = u_2 + \frac{\beta\gamma}{\beta+\gamma} \frac{(\sigma_1 - 1)N_1}{2C_1}. \tag{6.23b}$$

The total revenue in this pricing scheme is 0, i.e., $R_2 = 0$.

6.4.3　Pricing scheme 3

As for the third pricing scheme, we need to satisfy the following criteria:

C3-3a. The revenue is maximized during the congested period.

C3-3b. The cost for each SOV should not be higher than the no-toll scenario.

C3-3C. HOVs will not receive any credit.

The maximum revenue for operators is reached when $\Phi_2(t) = \phi_2(t)$. Thus, the tolls for each non-paying and paying SOV are:

$$u_2 = \frac{\beta\gamma}{\beta+\gamma}\frac{N_3^*}{C_2},$$

(6.24a)

$$u_3 = \frac{\beta\gamma}{\beta+\gamma}(\frac{N_2}{C_2} - \frac{N_1 + N_3^*}{C_1}).$$

(6.24b)

The maximum revenue for operators is calculated by:

$$R_3 = \frac{\beta\gamma}{\beta+\gamma}N_3^*(\frac{2N_2 - N_3^*}{C_2} - \frac{N_1 + N_3^*}{C_1})$$

(6.25)

6.5 Case study

In this section, we set $N_1 = 2250$ veh, $N_2 = 7200$ veh, $\sigma_1 = 2$, $C_1 = 1800$ vph, $C_2 = 3600$ vph, $\alpha = \$30/\text{hr}$, $\beta = \$15/\text{hr}$, and $\gamma = \$60/\text{hr}$. The desired arrival time as $t^* = 0$.

6.5.1 Departure time choice at the equilibrium state

Figure 6.4 shows the solutions when the HOT lanes are only available for HOVs. The x-axis represents the departure time for all vehicles. In Figure 6.4a, the length of the congested period for the HOT lanes is 1.25 hr, and the earliest and latest departing times for HOVs are $t_1^s = -1$ hr and $t_1^e = 0.25$ h, respectively. The departure flow-rate is 3600 vph for $t \in [-1, -0.5]$ hr, and 600 vph for $t \in [-0.5, 0.25]$ hr. For other times during this study period, the departure flow-rate is 0. In Figure 6.4b, the length of the congested period for the GP lanes is 2 hr. The earliest and latest departing times are $t_1^s = -1.6$ hr and $t_1^e = 0.4$ hr. The departure flow-rate is 7200 vph for $t \in [-1.6, -0.8]$ hr, and 1200 vph for $t \in [-0.8, 0.4]$ hr. Comparing the results in Figure 6.4a and Figure 6.4b, we can see that the GP lanes has

a longer congested period than the HOT lanes.

(a)   (b)
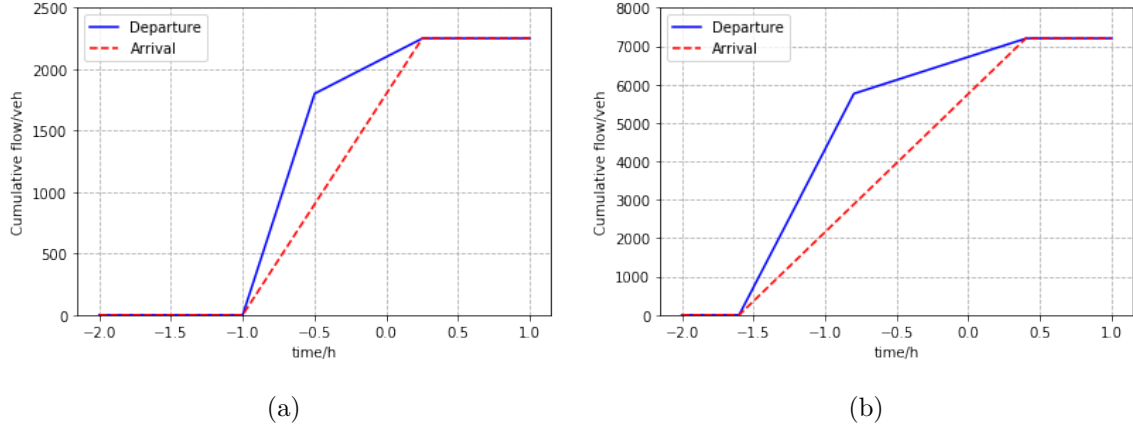
Figure 6.4: Cumulative $F(t)$ and $G(t)$ for each lane group in the no toll scenario: (a) HOT lanes; (b) GP lanes.

Figure 6.5 shows the cumulative departure and arrival curves for each lane group when the total travel time and scheduling cost is minimized. 150 SOVs will pay and switch to the HOT lanes. From Figure 6.5a, we can see that the earliest and latest departing times for HOVs and paying SOVs are $t_1^s = -16/15$ hr and $t_1^e = 4/15$ hr. The departure flow-rate is 3600 vph for $t \in [-16/15, -8/15]$ hr, and 600 vph for $t \in [-8/15, 4/15]$ hr. No vehicle departs beyond this time period. In Figure 6.4b, the earliest and latest departing times for non-paying SOVs are $t_1^s = -47/30$ hr and $t_1^e = 47/120$ hr. The departure flow-rate is 7200 vph for $t \in [-47/30, -47/60]$ hr, and 1200 vph for $t \in [-47/60, 47/120]$ hr. Comparing the results in Figure 6.4 and Figure 6.5, we find that the length of the congestion period increases by $\frac{1}{12}$ hr for the HOT lanes, but decreases by $\frac{1}{24}$ hr for the GP lanes in the existence of the pricing schemes.
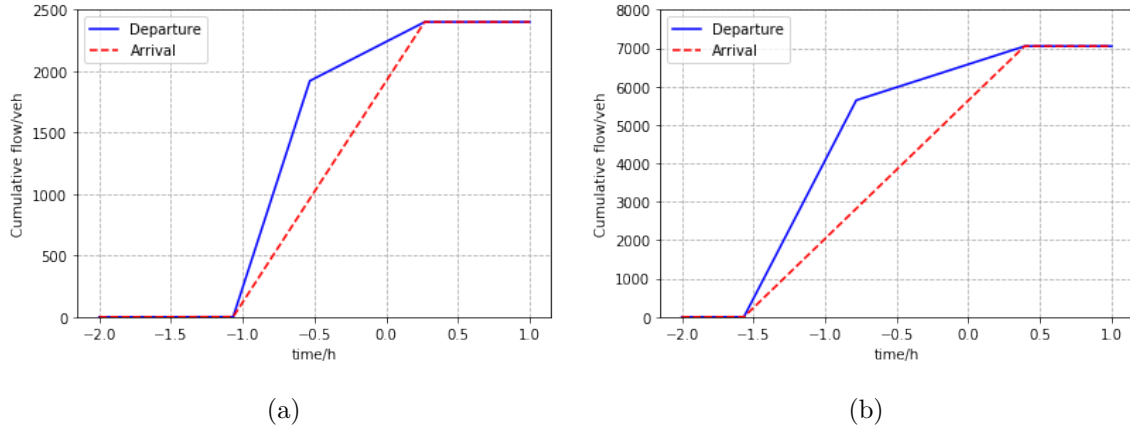
Figure 6.5: Cumulative $F(t)$ and $G(t)$ for each lane group when the total travel time and scheduling cost is minimized: (a) HOT lanes; (b) GP lanes.

6.5.2 Comparison of different pricing schemes

A comparison of the no-toll scenario and three pricing schemes is shown in Table 6.1. Five metrics are considered, including the total cost for a HOV, a non-paying SOV, and a paying SOV, the total travel time and scheduling cost (T), and the revenue for operators. The unit for all metrics is in \$. The total travel time and scheduling cost is calculated by (6.17). The revenues in scheme 1 and 3 are obtained from (6.21) and (6.25).

In the no-toll scenario, no vehicle pays a fee or receives a credit, the total travel time and scheduling cost is \$240, 300, which is the same as the total cost for all commuters; and the corresponding revenue is 0. For all three pricing schemes, the total travel time and scheduling cost is \$240, 075. In the first scheme, the cost for each HOV is \$16, and the cost for each SOV is \$23.5. Each paying SOV needs to pay \$7.5 to use the HOT lanes. In the second scheme, all SOVs have to pay \$0.15625 more than the first scheme. Operators can make use of this revenue to compensate the credits provided to HOVs. In the third scheme, the cost for each HOV increases by \$1, and each non-paying and a paying SOV needs to pay \$0.5 and \$8, respectively. The cost for each SOV is the same as the no-toll scenario. Operators

124

can gain a profit of $4,725$.

Table 6.1: System performance in different pricing schemes

	Φ_1	Φ_2	Φ_3	T	Revenue
No-toll	15	24	24	240,300	0
Scheme 1	16	23.5	23.5	240,075	1,125
Scheme 2	15	23.65625	23.65625	240,075	0
Scheme 3	16	24	24	240,075	4725

6.6 Comparison of pricing schemes with and without departure time choice

In most researches about pricing schemes for HOT lanes, the departure time choice behavior is not considered. In those studies, HOT lanes are "underutilized" when the demand of HOVs is lower than the capacity for the whole study period (Yin and Lou, 2009; Lou et al., 2011; Wang and Jin, 2017). When HOT lanes are uncongested but the parallel GP lanes are congested, a dynamic pricing scheme is required to maximize the flow-rate while maintaining the free-flow condition on HOT lanes. In those pricing schemes, the paying SOVs will not increase the HOVs' travel cost. Generally, a feedback pricing scheme is suggested to achieve the goal.

However, based on the daily observations, some commuters desire to depart either early or late to avoid congestion at the bottleneck. In this sense, it is more practical to consider both departure time and lane group choice when modeling commuters' behavior during the peak hours. When the departure time user equilibrium is reached, the sum of travel and scheduling costs is identical for vehicles in each type of vehicles. During peak periods, both

HOT and GP lanes are congested. We define the HOT lanes as "underutilized" when the congested period for the HOT lanes is shorter than that for the GP lanes. Any additional vehicle on the HOT lanes will increase the total cost for HOVs. Based on the analytical results, we find that when the drivers have the same VOT and penalty for early and late arrival, a simple flat toll scheme is sufficient to minimize the total travel time and scheduling cost. This type of problem can be treated as an optimization problem. Comparing with the dynamic pricing scheme, a flat toll is easier to implement and operate (Xiao et al., 2012).

6.7 Conclusion

In this study, we design pricing schemes for a freeway segment with HOT and GP lanes, considering commuters' departure time choice. When the departure time equilibrium is reached, the total cost for each vehicle on either HOT or GP lanes is the same (but the cost for a single HOV and SOV may not be identical). For each vehicle, the total cost is balanced by a high scheduling cost and a low queuing time cost, or vice versa. For the SOVs, they can adjust not only the lane group choice but also the departure time to balance their total cost. In the departure time equilibrium, the full-utilization principle defined in most previous HOT lane studies is not applicable. So, we are interested in exploring what types of pricing schemes should be applied to the HOT lanes.

We first analyze the equilibrium state with and without tolls, respectively. Then, we propose three criteria for the flat pricing schemes: (1) the total travel time and scheduling cost is minimized; and (2) the costs for each non-paying and paying SOV are the same. We further propose the third criterion based on the operators' requirement for revenue and the tolling constrains for certain type of vehicles. We show that the additional criterion leads to different pricing schemes. We provide a case study to present the departure time choice and compare the performance in different pricing schemes. We also compare pricing schemes with and

without consideration of departure time choice.

The following are some potential future research topics.

- In this study, we apply the homogeneous α-β-γ model in the cost function for SOVs and HOVs. We are interested in exploring the departure time and lane group choice for heterogeneous SOVs (Arnott et al., 1987).

- In the current study, the drivers can either arrive early or late. We also want to explore the case in which late arrival is not allowed. This type of problem can be formulated if we set an infinite penalty for the late arrival.

- In this study, we assume all commuters have the same ideal arrive time, t^*. We will consider the case in which t^* is distributed over a period, since commuters may live and work at a different distance from the bottleneck.

- In this study, paying SOVs and HOVs share the HOT lanes in the same period. In the follow-up study, we will design a new pricing scheme for a different time period. In the new pricing scheme, the paying SOVs will have access to the HOT lanes during the time periods when the HOT lanes are uncongested while the GP lanes are congested (i.e., the pricing scheme will be applied during $t \in [t_2^s, t_1^s] \cup [t_1^e, t_2^e]$). In this scenario, HOVs' travel costs are the same with and without pricing schemes, so they will not receive credits from operators.

- In this study, we show that the optimal demand of paying SOVs is determined by the average occupancy of HOVs. For various demand profiles, different average occupancy of HOVs should be considered to minimize the total cost for the system. In the future, we want to investigate how the minimum occupancy requirement of HOVs will impact the departure time and lane group choice. This will help determine the best policy for operating HOT lanes.

- In this study, we prove that constant tolls are sufficient to control the system when both the departure time and lane choice equilibrium is considered. This can be a potential scenario in which auctions can be applied to determine which SOVs can use the HOT lanes. In the follow-up study, we are interested in designing an auction-based pricing scheme for SOVs to achieve various goals, including maximizing social welfare and revenue for the operators.

Chapter 7

Conclusion

7.1 Summary

High-occupancy toll lanes have received more and more attention, especially by agencies in states like California, Texas, Washington, and Florida. However, a theoretic analysis of the system is limited in the literature. At the same time, current researches on HOT lanes operation have some drawbacks. For example, existing pricing strategies cannot guarantee that the closed-loop system converges to the optimal state, in which the HOT lanes capacity is fully utilized but there is no queue on the HOT lanes, and a well-behaved estimation and control method is quite challenging and still elusive.

In this dissertation, we study a simple freeway corridor with one origin and one destination. The corridor has three types of lanes: HOV lanes, HOT lanes, and GP lanes. There are two bottlenecks on the freeway, one for each lane group. HOVs can use the HOT lanes for free, but SOVs have to pay a price to use the HOT lanes. In this setup, we design pricing schemes that can achieve any two (but not all three) of the objectives: (1) maintaining a certain LOS on the HOT lanes, which includes keeping free-flow speed, keeping zero queue,

and keeping the speed above certain value; (2) improving the overall system performance (e.g., maximizing vehicle or person throughput, which is equivalent to minimize the travel time); and (3) maximizing revenue for the operators (Perez et al., 2012). We also consider the departure time choice problem in a separate chapter.

This dissertation is by the work of (Yin and Lou, 2009). They proposed two methods to provide dynamic tolling schemes for a HOT lane facility: the feedback method and the self-learning method. We prove that the feedback method cannot control the system when the freeway is congested. Meanwhile, we find that the self-learning method leads to a residual queue on the HOT lane. Combining the advantages of those methods, we provide a new control method to simultaneously estimate the average VOTs of SOVs and determining the dynamic price for the HOT lanes. We analytically and numerically prove that the closed-loop system is stable and guaranteed to converge to the optimal state,in either Gaussian or exponential manners. The effect of the scale parameter in the logit model is examined as well.

After realizing the limitations of the logit model, we propose a vehicle-based user equilibrium principle to capture the lane choice of SOVs. Based on the characteristics of those two models, we derive a general lane group choice model. We analytically solve the optimal dynamic prices with constant demands of HOVs and SOVs and obtain an insight regarding the dynamic price and then develop a feedback controller to determine the dynamic prices without knowing SOVs lane choice models, but to satisfy the two control objectives: to maximize the flow-rate but not to form a queue on the HOT lanes. We also show that if the type of the lane choice model is given, we can estimate the distribution of VOTs of the SOVs.

The consistency of the public and private operator's interest is critical in policy implications. It is claimed that revenue maximization should generally coincide with the optimization of freeway performances, such as maximizing overall travel-time savings or throughput (Perez

and Sciara, 2003). We propose an optimal control approach to get the dynamic prices that maximize the revenue for operators. We find that operators need to make different strategies based on the traffic demand. If the demand of HOVs is low, in order to maximize the revenue, operators should set a higher price to make the HOT lane underutilized. If the demand of HOVs is high, operators need to set a lower price to attract more SOVs to maximize revenue. In this case, both lanes will be congested. On the other hand, if they want to minimize total delay while keeping free-flow speed on the HOT lane, they need to sacrifice some revenue.

It has long been known that drivers' departure time choice behavior is one fundamental cause of congestion. If a driver departs early/late, s/he will experience lower queuing cost but higher scheduling cost; If departing close to the desire time, s/he will have more queuing cost but less scheduling cost. When the departure time user equilibrium is reached, the cost for vehicles on the same lane group is identical. In the presence of HOT lanes, SOVs can balance their total cost by adjusting their departure time and the lane group choice. In the last part of this dissertation, we propose pricing schemes that consider both lane choice and departure time choice. It turns out that flat (instead of dynamic) pricing schemes are able to meet the following three constraints: (1) the total travel time and scheduling cost is minimized; and (2) the costs for each non-paying and paying SOV are the same. We show that different revenue requirements and tolling constrains for certain type of vehicles can lead to different pricing schemes. We also compare the pricing schemes for HOT lanes with and without considering departure time choice.

7.2 Future research topics

One limitation of this study is the simplification of the road geometry. We only considered a freeway segment with one entrance to HOT lanes. It would be preferred if the current setup can be extended to HOT lanes with multiple entrances and multiple exits. And the

131

lane changing behavior at the weaving zones should be considered as well.

We show the price should be constant if the departure time user equilibrium is reached. This is a good starting point for investigating the auction-based pricing for HOT lanes. More specifically, each bidder uses a web-based platform to submit his or her bid for one permit to the operators. The operators then order the bids from highest to lowest. Suppose that the cap of the permits is N. Those N highest bidders will be the winners and pay the $(N+1)$ th highest bid for one permit. This mechanism is a special case of the well-known Vickrey-Clarke-Groves (VCG) mechanism, which has been proved to be incentive compatible (Parkes and Shneidman, 2004). An optimization problem can be formulated to find the number of permits needed to maximize the social welfare.

In Chapter 6, we propose and compare two definitions for "underutilized" HOT lanes: (1) the demand of HOVs is less than the capacity of the HOT lanes; and (2) the congestion period on the HOT lanes is shorter than that on the GP lanes. In the future study, we would like to conduct empirical studies with the data from SR-91 Express Lanes to check which definition is more realistic.

In Chapter 6, SOVs choose their departure time based on the flat tolls to reach an equilibrium state. We are interested in what departure time choice for SOVs would be if dynamic pricing schemes are applied. Will a stable day-to-day departure time choice exist?

Further research can also extend to a framework that includes car-sharing and ride-sharing behaviors when the lane choice and departure time are considered. We are interested in how pricing schemes change the mode choice of current SOV drivers. Will more SOVs choose to pay the congestion fees? Or will more SOVs want to share vehicles with other people?

Bibliography

Agnew, C.E., 1977. The theory of congestion tolls. Journal of Regional Science 17, 381–393.

Arnott, R., De Palma, A., Lindsey, R., 1987. Schedule delay and departure time decisions with heterogeneous commuters. University of Alberta, Department of Economics.

Arnott, R., De Palma, A., Lindsey, R., 1990a. Economics of a bottleneck. Journal of urban economics 27, 111–130.

Arnott, R., de Palma, A., Lindsey, R., 1990b. Departure time and route choice for the morning commute. Transportation Research Part B: Methodological 24, 209–228.

Aström, K.J., Murray, R.M., 2010. Feedback systems: an introduction for scientists and engineers. Princeton university press.

Bitran, G., Caldentey, R., 2003. An overview of pricing models for revenue management. Manufacturing & Service Operations Management 5, 203–229.

Boyles, S.D., Gardner, L.M., Bar-Gera, H., 2015. Incorporating departure time choice into high-occupancy/toll (hot) algorithm evaluation. Transportation Research Procedia 9, 90–105.

Brent, D.A., Gross, A., 2018. Dynamic road pricing and the value of time and reliability. Journal of Regional Science 58, 330–349.

Brownstone, D., Ghosh, A., Golob, T.F., Kazimi, C., Van Amelsfort, D., 2003. Drivers' willingness-to-pay to reduce travel time: evidence from the san diego i-15 congestion pricing project. Transportation Research Part A: Policy and Practice 37, 373–387.

Burris, M.W., Xu, L., 2006. Potential single-occupancy vehicle demand for high-occupancy vehicle lanes: Results from stated-preference survey of travelers in high-occupancy toll corridors. Transportation Research Record 1960, 108–118.

California Department of Transportation, 2019. Performance measurement system (pems). `http://http://pems.dot.ca.gov/`.

Campbell, R., 2016. Estimating drivers willingness to pay by using empirical data from a variably priced freeway facility. Transportation Research Record 2554, 1–9.

Chang, M.S., Hsueh, C.F., 2006. A dynamic road pricing model for freeway electronic toll collection systems under build–operate–transfer arrangements. Transportation Planning and Technology 29, 91–104.

Cheng, D., Ishak, S., 2015. Maximizing toll revenue and level of service on managed lanes with a dynamic feedback-control toll pricing strategy. Canadian Journal of Civil Engineering 43, 18–27.

Chou, F.S., Parlar, M., 2006. Optimal control of a revenue management system with dynamic pricing facing linear demand. Optimal Control Applications and Methods 27, 323–347.

Chung, C.L., Recker, W., 2011. State-of-the-art assessment of toll rates for high-occupancy and toll lanes. Technical Report.

Dafermos, S.C., 1972. The traffic assignment problem for multiclass-user transportation networks. Transportation science 6, 73–87.

Daganzo, C.F., 1994. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. Transportation Research Part B: Methodological 28, 269–287.

De Palma, A., Lindsey, R., 2000. Private toll roads: Competition under various ownership regimes. The Annals of Regional Science 34, 13–35.

Dorogush, E.G., Kurzhanskiy, A.A., 2015. Modeling toll lanes and dynamic pricing control. arXiv preprint arXiv:1505.00506 .

Emmerink, R., Nijkamp, P., Rietveld, P., 1995. Is congestion pricing a first-best strategy in transport policy? a critical review of arguments. Environment and Planning B: Planning and Design 22, 581–602.

Fuhs, C., Obenberger, J., 2002. Development of high-occupancy vehicle facilities: Review of national trends. Transportation Research Record 1781, 1–9.

Gardner, L.M., Bar-Gera, H., Boyles, S.D., 2013. Development and comparison of choice models and tolling schemes for high-occupancy/toll (hot) facilities. Transportation Research Part B: Methodological 55, 142–153.

Göçmen, C., Phillips, R., van Ryzin, G., 2015. Revenue maximizing dynamic tolls for managed lanes: A simulation study.

Habib, T.M.A., 2013. Simultaneous spacecraft orbit estimation and control based on gps measurements via extended kalman filter. The Egyptian Journal of Remote Sensing and Space Science 16, 11–16.

Hendrickson, C., Kocur, G., 1981. Schedule delay and departure time decisions in a deterministic model. Transportation science 15, 62–77.

Ieromonachou, P., Potter, S., Warren, J.P., 2006. Norway's urban toll rings: Evolving towards congestion charging? Transport Policy 13, 367–378.

Jang, K., Chung, K., Ragland, D., Chan, C.Y., 2009. Safety performance of high-occupancy-vehicle facilities: Evaluation of hov lane configurations in california. Transportation Research Record: Journal of the Transportation Research Board 2099, 132–140.

Jang, K., Chung, K., Yeo, H., 2014. A dynamic pricing strategy for high occupancy toll lanes. Transportation Research Part A: Policy and Practice 67, 69–80.

Janson, M., Levinson, D., 2014. Hot or not: Driver elasticity to price on the mnpass hot lanes. Research in Transportation Economics 44, 21–32.

Jin, W.L., 2013. A multi-commodity lighthill-whitham-richards model of lane-changing traffic flow. Procedia-Social and Behavioral Sciences 80, 658–677.

Jin, W.L., 2015. Point queue models: A unified approach. Transportation Research Part B: Methodological 77, 1–16.

Jin, W.L., Gan, Q.J., Lebacque, J.P., 2015. A kinematic wave theory of capacity drop. Transportation Research Part B: Methodological 81, 316–329.

Jin, W.L., Wang, X., Lou, Y., 2018. Control and estimation of traffic systems with high-occupancy-toll lanes, in: Proceedings of the 97th Annual Meeting of the Transportation Research Board.

Joksimovic, D., Bliemer, M.C., Bovy, P.H., 2005. Optimal toll design problem in dynamic traffic networks with joint route and departure time choice. Transportation Research Record 1923, 61–72.

Khalil, H.K., 2002. Nonlinear systems. Upper Saddle River .

Kirk, D.E., 2012. Optimal control theory: an introduction. Courier Corporation.

Knight, F.H., 1924. Some fallacies in the interpretation of social cost. The Quarterly Journal of Economics 38, 582–606.

Kuwahara, M., 1990. Equilibrium queueing patterns at a two-tandem bottleneck during the morning peak. Transportation science 24, 217–229.

Kwon, J., Varaiya, P., 2008. Effectiveness of california's high occupancy vehicle (hov) system. Transportation Research Part C: Emerging Technologies 16, 98–115.

Lam, T.C., Small, K.A., 2001. The value of time and reliability: measurement from a value pricing experiment. Transportation Research Part E: Logistics and Transportation Review 37, 231–251.

Laval, J.A., Cho, H.W., Muñoz, J.C., Yin, Y., 2015. Real-time congestion pricing strategies for toll facilities. Transportation Research Part B: Methodological 71, 19–31.

Levy-Lambert, H., 1968. Tarification des services à qualité variable–application aux péages de circulation. Econometrica: Journal of the Econometric Society .

Liberzon, D., 2003. Switching in systems and control. Springer Science & Business Media.

Liu, H.X., He, X., Recker, W., 2007. Estimation of the time-dependency of values of travel time and its reliability from loop detector data. Transportation Research Part B: Methodological 41, 448–461.

Liu, X., Zhang, G., Lao, Y., Wang, Y., 2011. Quantifying the attractiveness of high-occupancy toll lanes with traffic sensor data under various traffic conditions. Transportation Research Record 2229, 102–109.

Lou, Y., Yin, Y., Laval, J.A., 2011. Optimal dynamic pricing strategies for high-occupancy/toll lanes. Transportation Research Part C: Emerging Technologies 19, 64–74.

Mahmassani, H., Herman, R., 1984. Dynamic user equilibrium departure time and route choice on idealized traffic arterials. Transportation Science 18, 362–384.

Marchand, M., 1968. A note on optimal tolls in an imperfect environment. Econometrica: Journal of the Econometric Society , 575–581.

May, A.D., Milne, D.S., 2000. Effects of alternative road pricing systems on network performance. Transportation Research Part A: Policy and Practice 34, 407–436.

Michalaka, D., Lou, Y., Yin, Y., 2011. Proactive and robust dynamic pricing strategies for high-occupancy-toll (hot) lanes, in: 90th Annual Meeting of the Transportation Research Board, Washington, DC.

Nikolic, G., Pringle, R., Jacob, C., Mendonca, N., Bekkers, M., Torday, A., Rinelli, P., 2014. Dynamic tolling of hot lanes through simulation of expected traffic conditions, in: Transportation 2014: Past, Present, Future-2014 Conference and Exhibition of the Transportation Association of Canada.

Nikoobin, A., Moradi, M., 2017. Indirect solution of optimal control problems with state variable inequality constraints: finite difference approximation. Robotica 35, 50–72.

OCTA, 2003. 91 express lanes toll policy. `http://https://www.91expresslanes.com/wp-content/uploads/2014/04/TollPolicy.pdf`.

Ohnishi, K., Matsui, N., Hori, Y., 1994. Estimation, identification, and sensorless control in motion control system. Proceedings of the IEEE 82, 1253–1265.

de Palma, A., Lindsey, R., 2011. Traffic congestion pricing methodologies and technologies. Transportation Research Part C: Emerging Technologies 19, 1377–1399.

Pandey, V., Boyles, S.D., 2018. Dynamic pricing for managed lanes with multiple entrances and exits. Transportation Research Part C: Emerging Technologies 96, 304–320.

Pandey, V., Boyles, S.D., 2019. Comparing route choice models for managed lane networks with multiple entrances and exits. Transportation Research Record , 0361198119848706.

Papageorgiou, M., Hadj-Salem, H., Blosseville, J.M., et al., 1991. Alinea: A local feedback control law for on-ramp metering. Transportation Research Record 1320, 58–67.

Parkes, D.C., Shneidman, J., 2004. Distributed implementations of vickrey-clarke-groves mechanisms, in: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1, IEEE Computer Society. pp. 261–268.

Perez, B.G., Fuhs, C., Gants, C., Giordano, R., Ungemah, D.H., 2012. Priced managed lane guide. Technical Report. United States. Federal Highway Administration.

Perez, B.G., Sciara, G.C., 2003. A guide for HOT lane development. US Department of Transportation, Federal Highway Administration.

Pigou, A.C., 1920. The economics of welfare. Macmillan and Co., London.

Schrank, D., Eisele, B., Lomax, T., Bak, J., 2015. 2015 urban mobility scorecard .

Small, K.A., 1982. The scheduling of consumer activities: work trips. The American Economic Review 72, 467–479.

Small, K.A., 2015. The bottleneck model: An assessment and interpretation. Economics of Transportation 4, 110–117.

Sorensen, P., Wachs, M., Daehner, E.M., Kofner, A., Ecola, L., Hanson, M., Yoh, A., Light, T., Griffin, J., 2008. Reducing Traffic Congestion in Los Angeles. Technical Report.

Sullivan, Edward, 2000. Continuation study to evaluate the impacts of the sr-91 value-priced express lanes. final report. http://ceenve3.civeng.calpoly.edu/sullivan/SR91/final_rpt/FinalRep2000.pdf.

Supernak, J., Steffey, D., Kaschade, C., 2003. Dynamic value pricing as instrument for better utilization of high-occupancy toll lanes: San diego i-15 case. Transportation Research Record: Journal of the Transportation Research Board 1839, 55–64.

Szeto, W., Lo, H.K., 2004. A cell-based simultaneous route and departure time choice model with elastic demand. Transportation Research Part B: Methodological 38, 593–612.

Tan, Z., Gao, H.O., 2018. Hybrid model predictive control based dynamic pricing of managed lanes with multiple accesses. Transportation Research Part B: Methodological 112, 113–131.

Taylor, L., 1970. The existence of optimal distributed lags. The Review of Economic Studies 37, 95–106.

Train, K.E., 2009. Discrete choice methods with simulation. Cambridge university press.

TRB Managed Lanes Committee, 2019. Projects database. https://managedlanes.wordpress.com/2017/07/07/projects-database.

Tseng, Y.Y., Verhoef, E.T., 2008. Value of time by time of day: A stated-preference study. Transportation Research Part B: Methodological 42, 607–618.

U.S. Department of Transportation, 2016. Revised departmental guidance on valuation of travel time in economic analysis. `https://www.transportation.gov/office-policy/transportation-policy/revised-departmental-guidance-valuation-travel-time-economic`.

Verhoef, E., Nijkamp, P., Rietveld, P., 1996a. Second-best congestion pricing: the case of an untolled alternative. Journal of Urban Economics 40, 279–302.

Verhoef, E., Nijkamp, P., Rietveld, P., 1996b. Second-best congestion pricing: the case of an untolled alternative. Journal of Urban Economics 40, 279–302.

Vickrey, W.S., 1969. Congestion theory and transport investment. The American Economic Review 59, 251–260.

Viti, F., Catalano, S.F., Li, M., Lindveld, C., van Zuylen, H.J., 2003. An optimization problem with dynamic route-departure time choice and pricing, in: Proceedings of the 82nd Annual Meeting of the Transportation Research Board.

Walters, A.A., 1961. The theory and measurement of private and social cost of highway congestion. Econometrica: Journal of the Econometric Society , 676–699.

Wang, X., Jin, W.L., 2017. A new method to estimate value of time for high-occupancy-toll lane operation, in: Proceedings of the 96th Annual Meeting of the Transportation Research Board.

Wardrop, J.G., 1952. Some theoretical aspects of road traffic research. Proceedings of the Institution of Civil Engineers 1, 325–378.

Wie, B.W., Tobin, R.L., Friesz, T.L., Bernstein, D., 1995. A discrete time, nested cost operator approach to the dynamic network user equilibrium problem. Transportation Science 29, 79–92.

Xiao, F., Shen, W., Zhang, H.M., 2012. The morning commute under flat toll and tactical waiting. Transportation Research Part B: Methodological 46, 1346–1359.

Yang, H., Huang, H.J., 1997. Analysis of the time-varying pricing of a bottleneck with elastic demand using optimal control theory. Transportation Research Part B: Methodological 31, 425–440.

Yang, H., Huang, H.J., 1998. Principle of marginal-cost pricing: how does it work in a general road network? Transportation Research Part A: Policy and Practice 32, 45–54.

Yang, H., Huang, H.J., 2005. Mathematical and economic theory of road pricing.

Yang, H., Meng, Q., 1998. Departure time, route choice and congestion toll in a queuing network with elastic demand. Transportation Research Part B: Methodological 32, 247–260.

Yang, L., 2012. Stochastic traffic flow modeling and optimal congestion pricing .

Yin, Y., Lou, Y., 2009. Dynamic tolling strategies for managed lanes. Journal of Transportation Engineering 135, 45–52.

Zhang, G., Wang, Y., Wei, H., Yi, P., 2008. A feedback-based dynamic tolling algorithm for high-occupancy toll lane operations. Transportation Research Record: Journal of the Transportation Research Board 2065, 54–63.

Zhang, J., Chang, H., Ioannou, P.A., 2006. A simple roadway control system for freeway traffic, in: 2006 American Control Conference, IEEE. pp. 4900–4905.