## Title

Freeway Traffic Parameter and State Estimation with Eulerian and Lagrangian Data

## Permalink

https://escholarship.org/uc/item/521733s2

## Author

Sun, Zhe

## Publication Date

2016

## Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-ShareAlike License, availalbe at https://creativecommons.org/licenses/by-nc-sa/4.0

Peer reviewed|Thesis/dissertation

# UNIVERSITY OF CALIFORNIA, IRVINE 

# Freeway Traffic Parameter and State Estimation with Eulerian and Lagrangian Data DISSERTATION 

submitted in partial satisfaction of the requirements
for the degree of

## DOCTOR OF PHILOSOPHY

in Civil and Environmental Engineering
by

Zhe Sun

Dissertation Committee:
Professor Wenlong Jin, Chair
Professor R. Jayakrishnan
Professor Will W. Reckor
(c) 2016 Zhe Sun

## DEDICATION

To my family

## TABLE OF CONTENTS

Page
LIST OF FIGURES ..... iv
LIST OF TABLES ..... v
ACKNOWLEDGMENTS ..... vi
CURRICULUM VITAE ..... vii
ABSTRACT OF THE DISSERTATION ..... ix
1 Introduction ..... 1
1.1 Research Background ..... 1
1.2 Research Objective ..... 2
1.3 Research Outline ..... 4
2 Literature Review ..... 7
2.1 Traffic sensor data ..... 7
2.2 Kinematic wave theory of traffic flow ..... 11
2.3 Traffic parameter and state estimation ..... 14
3 Network Sensor Health Problem ..... 18
3.1 Introduction ..... 18
3.2 Formulation ..... 20
3.2.1 The Node-based Flow Conservation ..... 21
3.2.2 Sensor Health Index ..... 22
3.3 Solution Algorithm ..... 28
3.3.1 The brute-force enumeration ..... 28
3.3.2 The greedy search algorithm ..... 29
3.4 A Real World Example ..... 35
3.5 Summary ..... 38
4 Traffic Estimation with Vehicle Reidentification Data ..... 39
4.1 Introduction ..... 39
4.2 The LoopReid estimation method ..... 41
4.2.1 An optimization problem in initial states and parameters ..... 42
4.2.2 Estimation of traffic state ..... 45
4.3 Solution of the optimization problem ..... 47
4.3.1 A decoupling method ..... 47
4.3.2 Potential issue with near-stationary traffic ..... 49
4.4 Applications to NGSIM I101 dataset ..... 51
4.4.1 Data Preparation ..... 52
4.4.2 Estimation of initial states and model parameters ..... 54
4.4.3 Validation with respect to the average density and upstream flow-rate ..... 56
4.4.4 A comparison with the STD method ..... 58
4.5 Summary ..... 61
5 Traffic Estimation with Complete Trajectory Data ..... 62
5.1 Introduction ..... 62
5.2 A hybrid kinematic wave model in Lagrangian coordinates under Eule-63
5.3 The LoopCT estimation method ..... 73
5.3.1 An optimization problem in initial states and parameters ..... 74
5.3.2 Estimation of vehicle trajectories ..... 77
5.3.3 A comparison between LoopReid method and LoopCT method ..... 79
5.4 Solution and Optimization Method ..... 80
5.4.1 A decoupling method with predetermined free-flow speed ..... 80
5.4.2 Potential issue with near-stationary state traffic ..... 84
5.5 Application to NGSIM I80 dataset ..... 85
5.5.1 Data Preparation ..... 86
5.5.2 Estimation of initial states and model parameters ..... 87
5.5.3 Validation using average density and cumulative count contour ..... 89
5.5.4 Vehicle trajectory estimation ..... 92
5.6 Summary ..... 93
6 Conclusion ..... 95
6.1 Summary ..... 95
6.2 Future research directions ..... 97
Bibliography ..... 99

## LIST OF FIGURES

## Page

1.1 Dissertation framework ..... 4
2.1 Configuration of California loop detector network in District 7 ..... 9
2.2 Triangular fundamental diagram ..... 12
2.3 Newell's model in the U-shape spatial-temportal domain ..... 13
2.4 A linear discrete Kalman filter ..... 16
3.1 Illustration of relations between link sets ..... 23
3.2 Sample Network ..... 25
3.3 Calculation of sensor health index ..... 27
3.4 Swap operation ..... 32
3.5 Solution procedure for the sample network ..... 34
3.6 The SR-91 east network ..... 35
3.7 Lattice of solutions ..... 36
3.8 Health Index (Greedy Search) v.s. PeMS ..... 37
3.9 Health Index (Brute-force) v.s. PeMS ..... 38
4.1 Illustration of a vehicle trajectory and kinematic waves on a road segment ..... 41
4.2 Illustration of the Study Site ..... 51
4.3 Piecewise linear interpolation of cumulative flows ..... 53
4.4 Plot of the objective function ..... 55
4.5 Estimated upstream flow-rate v.s. observation ..... 58
4.6 Impact of Flow Counting Interval and CMR on Average Density Esti- ..... 60
5.1 Illustration of relation of $N_{1}(t, x)$ and $N_{2}(t, x)$ ..... 66
5.2 Representation of $X_{1}(t, i)$ and $X_{2}(t, i)$ ..... 71
5.3 Illustration of the LoopCT estimation method ..... 75
5.4 Vehicle trajectory estimation ..... 78
5.5 Typical vehicle trajectories ..... 82
5.6 Stationary state with constant queue ..... 85
5.7 Illustration of I-80 study site ..... 86
5.8 Objective function plot with trace ..... 89
5.9 Absolute difference between the estimated and observed cumulative90
5.10 Observed density contour plot . . . . . . . . . . . . . . . . . . . . . . 91
5.11 Estimation v.s. observation, Vehicle \#3164. . . . . . . . . . . . . . . 92

## LIST OF TABLES

Page
3.1 SSE for all base sets ..... 27
4.1 Verification results ..... 54
4.2 MAPE of estimated average density and upstream flow-rate ..... 59
5.1 Parameter estimation ..... 88
5.2 MAPE/MAE of estimated average density and cumulative contour . 90
5.3 Trajectory estimation error ..... 93

## ACKNOWLEDGMENTS

I would like to express my deepest gratitude to Professor Wenlong Jin for his advice during my doctoral study for the past five years. As my advisor, he has constantly encouraged me to remain focused on my goals. His observations and comments helped me to establish the direction of the research and to move forward with investigation in depth. The joy and enthusiasm he has for research is contagious and motivational for me, even during the tough times in my Ph.D. journey. He set an example for me and taught me the power of introspection, which is key to become a successful researcher.

Appreciation is also extended to Professor Wilfred W. Reckor and Professor R. Jayakrishnan, my committee members, for their time and valuable advices. I also would like to thank Professor Stephen G. Ritchie for his discussions on the traffic estimation method, Professor Michael G. McNally, Professor Jean-Daniel Saphores for their teaching and encouragement, and Professor ManWo Ng for his advices on the network sensor health problem.

I am deeply indebted to Ms. Ziggy Bates who offered me the opportunity to work as her student assistant. I also thank Ms. Anne Marie DeFeo, Ms. Kathy Riley, Ms. Myra Radlow and all ITS staff for their kind assistance and friendship.

I enjoyed working with my colleagues, Dr. Hao Yang and Dr. Qijian Gan, in Professor Jin's research team and I have learned so much from them. I warmly thank my officemates, Suman Mitran, Anupam Srivastava, and Ke Wang (Kevin) for their selfless help and support, especially when I felt frustrated and lack of motivation. My special thanks go to Adrian Rey, whose achievements in vehicle trajectory estimation greatly inspired my own research.

I would like to express my gratitude to all the ITS friends including Qinglong Yan, Felipe Augusto, Shizhe Shen (Tony), Yue Zhou, Yue Sun (Ethan), Xuting Wang, Jiangbo Yu (Gabriel), Xiaoxia Shi (Shia), Harya Dillon (Koko), Kyungsoo Jeong, Kyung Hyun (Kate), Wan-Tzu Lo (Ashley), and Neda Masoud.

Last but not least, I would like to take this opportunity to express my gratitude and thanks to my family. My hardworking parents, Mr. Xiaochun Sun and Ms. Renli Liu, provided unconditional love and support which I can never repay. I feel very fortunate to meet my wife, Zheng Guo, who continues to stand behind me with endless love. Her support and encouragement was in the end what made this dissertation possible.

## CURRICULUM VITAE

Zhe Sun

## EDUCATION

Ph.D. in Transportation System Engineering<br>2016<br>University of California, Irvine<br>Irvine, California<br>M.S. in Statistics<br>2014<br>University of California, Irvine<br>Irvine, California<br>\(\begin{array}{lr}M.S. in Transportation System Engineering \& 2011<br>University of California, Irvine \& Irvine, California\end{array}\)<br>B.S. in Transportation Engineering<br>2010<br>Tongji University<br>Shanghai, China

## SELECTED HONORS AND AWARDS

## Honorable Mention Award

Transportation Research Board 2015 Data Competition
Graduate Research Fellowship
2011
University of California, Irvine
National Scholarship for Outstanding Undergraduate Student
2007
Tongji University, Shanghai

## REFERRED JOURNAL PUBLICATION

Sun, Z., Jin, W. L., \& Ritchie, S. G. (2016). Simultaneous State and Parameter Estimation in Newell's Simplified Kinematic Wave Model with Heterogeneous Data. Submitted to Transportation Research Part B.

Sun, Z., Jin, W. L., \& Ng, M. (2016). Network Sensor Health Problem. Submitted to Transportation Research Part C.

Sun, Z., \& Yang, C. (2011). Development of Extreme Value Distribution Method for Estimation of Expressway Capacity. Transportation Research Record: Journal of the Transportation Research Board, (2260), 133-139.

REFERRED CONFERENCE PUBLICATION

Yan, Q., Sun, Z., Gan, Q., \& Jin, W. L. (2016). Automatic near-stationary traffic state identification based on PELT changepoint detection. Presented at the 95th Annual Meeting of the Transportation Research Board, Washington, D.C.

Sun, Z., \& Ombao, H. (2015). Short-Term Traffic Flow Prediction with Mixture Autoregressive Model. Presented at the 94th Annual Meeting of the Transportation Research Board, Washington, D.C.

Sun, Z., Jin, W. L., \& Ritchie, S. G. (2015). Simultaneous State and Parameter Estimation in Newell's Simplified Kinematic Wave Model with Heterogeneous Data. Presented at the 94th Annual Meeting of the Transportation Research Board, Washington, D.C.

Sun, Z., Jin, W. L., \& Ng, M. (2014). Quality Diagnosis of Sensor Flow Data on a Freeway Network. Presented at the 93rd Annual Meeting of the Transportation Research Board, Washington, D.C.

Yang, H., Andres, L., Sun, Z., Gan, Q., \& Jin, W. L. (2014). A Field Test of a Dynamic Green Driving Strategy Based on Inter-Vehicle Communications. Presented at the 93rd Annual Meeting of the Transportation Research Board, Washington, D.C.

Gan, Q., Chu, L., Sun, Z., Yang, M. H., \& Jin, W. L. (2013), Calibration of the cell transmission model for a freeway network. Presented at the 92nd Annual Meeting of the Transportation Research Board, Washington, D.C.

Jin, W. L., Kwan, C., Sun, Z., Yang, H., \& Gan, Q. (2012). SPIVC: smart-phonebased intervehicle communication system. Presented at the 91st Annual Meeting of the Transportation Research Board, Washington, D.C.

## EXPERIENCE

Graduate Student Researcher
University of California, Irvine
2011--2016
Irvine, California

Service Development Intern
2013--2014
Long Beach Transit
Long Beach, California

## PROFESSIONAL SERVICE

Member, Chinese Overseas Transportation Association 2013 to Present
Member, Transportation Research Board
2010 to Present

## ABSTRACT OF THE DISSERTATION

Freeway Traffic Parameter and State Estimation with Eulerian and Lagrangian Data By

Zhe Sun

Doctor of Philosophy in Civil and Environmental Engineering
University of California, Irvine, 2016
Professor Wenlong Jin, Chair

The purpose of this study is to develop a traffic estimation framework which combines different data sources to better reconstruct the traffic states on the freeways. The framework combines both traffic parameter and state estimation in the same work flow, which resolves the inconsistency issue of most existing traffic state estimation methods.

To examine the quality of the traffic sensor data, the study starts with proposing the network sensor health problem (NSHP). The optimal set of sensors is selected from all sensors such that the violation of flow conservation is minimized. The health index for individual detector is then calculated based on the solutions. We also developed a tailored greedy search algorithm to find the solutions effectively. The proposed method is tested using the loop detector data from PeMS on a stretch of the SR-91 freeway. We compared the results with PeMS health status and found considerable level of consistency.

Two different traffic state estimation methods are proposed based on the data availability and traffic states. The LoopReid method is derived from the Newell's simplified kinematic wave model by assuming the whole road segment is fully congested. We formulate a least square optimization problem to find the initial states and traffic
parameters based on the first-in-first-out principle and the congested part of the Newell's model. While developing the LoopCT method, we derived a counterpart of the Newell's kinematic wave model in the Lagrangian coordinates under Eulerian boundary conditions. This model also leads to a new method to estimate vehicle trajectories within a road segment. We formulate a least square optimization problem in initial states and traffic parameters which works for mixed traffic states. The two estimation methods turned out to be highly related and the LoopCT method degenerates to the LoopReid method when the traffic is fully congested. The two methods are validated using two datasets from the NGSIM project. Both methods achieved considerable level of accuracy at reconstructing the traffic states and parameters.

## Chapter 1

## Introduction

### 1.1 Research Background

The traffic state is an important indicator of a traffic system's performance. However, the observed traffic data can only cover a limited number of locations and times due to limited deployment of sensing infrastructure. At the same time, the development of different traffic sensor technologies is changing the form of the traffic data fundamentally. In general, we have too few data, yet too many types of data. It would be ideal to combine different data sources to augment the data availability.

The traffic estimation with data fusion is becoming popular in recent years with big data boom. However, most of the existing studies neglect the estimation of traffic parameters and initial states, which may result in suboptimal results due to the use of unsuitable parameters and initial states. This dissertation proposes an estimation framework which attaches importance to both traffic parameter and state estimation with data fusion.

### 1.2 Research Objective

This study aims to propose a traffic estimation framework which simultaneously estimate traffic parameters and states on a homogeneous freeway link based on data fusion. To achieve this goal, the detailed objectives are described as follows:

## - To design a network-based sensor health evaluation method.

The performance of a traffic estimation method depends on the quality of the input data. Most existing sensor health evaluation methods are local methods which focus on the data yielded by individual sensors. The network-level flow conservation information should also be examined to ensure the consistency among sensors.

## - To develop a traffic estimation method based on both Lagrangian and

 Eulerian sensor dataThe Eulerian and Lagrangian data are fundamentally different. In this research, we propose an estimation method based on Newell's simplified kinematic wave model to incorporate both types of data. The LoopReid model combines loop detector data and vehicle reidentification data while the LoopCT model combines loop detector data with complete vehicle trajectory data.

## - To prove the single transition theorem within a homogeneous link

 The single transition theorem states that the vehicle will experience at most one transition from free-flow traffic to congested traffic while traveling through the road segment. The vehicle travels with the free-flow speed before the transition and reduces speed afterward. The proof of this theorem enables the estimation method to work with mixed traffic by separating the whole trajectory into two parts, each of which is either a free-flow or congested traffic state.- To apply the proposed traffic estimation method to the NGSIM data. The LoopReid method only works with congested traffic while the LoopCT method works with mixed traffic state. To evaluate the performance of the proposed methods, two NGSIM datasets are used. The LoopReid method is validated using the I101 dataset collected during evening peak when traffic is fully congested in the road segment. The LoopCT data is validated using the I80 dataset collected in the afternoon time when traffic congestion started to initiate.


### 1.3 Research Outline

The overall dissertation framework is described in Figure 1.1


Figure 1.1: Dissertation framework

Chapter 1 introduces the background and objectives of the dissertation. Chapter 2 reviews the literature regarding the three important components of the proposed estimation framework: traffic sensor data, traffic flow model, and the traffic estimation method.

Chapter 3 proposes the network sensor health problem (NSHP) and the corresponding sensor health evaluation method which takes network flow conservation into consideration to evaluate sensor health. From the network flow conservation principle, flows on non-base links can be derived from those on the base links. However, in reality, the network flow conservation principle can be violated due to the existence of unhealthy sensors. Thus we propose to identify optimal sensor sets by solving an optimization problem, in which we minimize the inconsistency between derived and observed link flows. We then define the health index of an individual sensor as the frequency that it appears in the optimal sets. The brute-force method can be used to find all solutions. We also developed a greedy search algorithm to find optimal sets effectively. The proposed method is applied to a road network with 30 links,
among which 18 links are monitored with loop detectors. Using traffic count data from the Caltrans Performance Measurement System (PeMS) database, the method yields health indices for all observed sensors and the results are quite consistent with PeMS health statuses.

Chapter 4 proposes the LoopReid traffic estimation method to simultaneously estimate model parameters and traffic states for a congested road segment based on Newell's simplified kinematic wave model (Newell, 1993). In many existing estimation methods, the model parameters and initial states have to be given, which limits the accuracy of the results as well as their transferability to different locations and times. Given both Eulerian traffic count data and Lagrangian vehicle reidentification data, we formulate a single optimization problem in terms of the initial number of vehicles and model parameters. Then we decouple the optimization problem such that the initial number of vehicles in the segment can be analytically solved in closed form, and the model parameters, including the jam density and the shock wave speed in congested traffic, can be computed with the Gauss-Newton method. Based on Newell's model, we can calculate individual vehicles' trajectories as well as the average densities, speeds, and flow-rates inside the road segment. We also theoretically show that the optimization problem can have multiple solutions under absolute stationary traffic conditions. The proposed method is applied to the I101 NGSIM dataset. We verify the validity of the method and show that this method yields better results in the estimation of average densities than the benchmark method.

Chapter 5 proposes the LoopCT traffic estimation method which can be considered as an extension of the LoopReid method based on complete vehicle trajectories. In order to derive the LoopCT method, we theoretically prove the single transition theorem and the counter part of the Newell's kinematic wave model in the Lagrangian coordinate under Eulerian boundary conditions. This model also leads to a new
method to estimate vehicle trajectories within a link given upstream/downstream flow counts and vehicle entrance/exit time. The estimation of the initial states and model parameters is formulated as an optimization problem. Similar to the LoopReid model, the initial states can be optimized analytically. By assuming the free-flow speed is known in advance, the optimization problem of traffic parameters is greatly simplified and can be estimated using the Gauss-Newton method. In theory, we show that there will be multiple solutions for traffic parameters under absolute stationary traffic conditions. We test the LoopCT method using the I80 NGSIM dataset with different market penetration rates and predetermined free-flow speeds. The average density, cumulative count contour, and the vehicle trajectories are estimated and compared with the observation.

Chapter 6 summarize the results and findings of this dissertation along with recommendations for future research.

## Chapter 2

## Literature Review

This chapter reviews the three elements of the traffic estimation framework: data, theory, and the method. Section 1 of this chapter reviews available data sources for traffic estimation with emphasis on existing methods to evaluate loop detector data quality. Section 2 of this chapter reviews the kinematic wave model of traffic flow with emphasis on Newell's simplified kinematic wave model. Section 3 reviews the existing traffic estimation models.

### 2.1 Traffic sensor data

Since the introduction of the first known vehicle sensor in 1928 at a signalized intersection, researchers have devoted significant efforts to create and improve systems that monitor vehicle presence and passage at critical locations on streets and freeways. The collected data are used to monitor traffic congestion and incidents on freeways, estimate travel times, and support decision making for transportation agencies. Many different types of traffic sensors, including inductive loop detectors, magnetic sensors,
video image detectors, and microwave detectors, have been developed.

In the study of fluid dynamics, the observation of fluid flow can be made in two ways: the Lagrangian measurements and the Eulerian measurements Batchelor, 2000). In the context of traffic flow, we can collect corresponding traffic data using these two measurements. The Eulerian data are collected at fixed locations and provides information regarding vehicles as they pass over them. The inductive loop detector system is the most commonly used sensing system to collect Eulerian data. These detectors measure traffic counts and occupancy, and aggregate them at certain sampling intervals (usually 30 seconds) for each lane. This type of data has been used for many traffic applications and studies, including traffic management and control and traffic state estimation.

The Lagrangian data are collected for individual vehicles. This type of data can be generated by

- vehicle reidentification system, which matches vehicles passing different locations (partial trajectory) (Sun et al., 1999). The vehicle reidentification system can be implemented based upon different sensing technologies, such as video cameras, AVI (automatic vehicle identification) tags, and loop detectors (Jeng, 2007). This type of data has been used for travel time estimation Coifman and Cassidy, 2002), performance evaluation (Jeng, 2007, Oh et al., 2005), and O/D trip estimation (Oh et al., 2002).
- vehicle tracking system, which provides sequences of location and time stamp of individual vehicles (complete trajectory). The vehicle tracking system commonly uses GPS technologies to locate the vehicles, but other sensing technologies such as RADAR (Aoude et al., 2011) and photos (Hoogendoorn et al., 2003) are also used.

Because of their high reliability under different weather conditions, the loop detectors have been widely deployed to provide uninterrupted traffic measurements, including occupancy and flow counts. In California, loop detectors are embedded in many freeway pavements, providing 30 -second and 5-minute occupancies and traffic counts data through the Caltrans Performance Measurement System (PeMS).

However, loop detector data can be corrupted by noises and errors, due to pavement/sawcut failures, intermittent communications, double counting of lane-changing vehicles, and so on (Coifman, 2006). According to PeMS, only $67 \%$ of the detectors are working properly in May, 2014 (Chen and Petty, 2001). Some districts (e.g. district 6 in Los Angeles County) have even lower proportions of working detectors. Thus, to accurately estimate congestion levels, incident locations, and travel times as well as to decide detectors that need to be maintained, the state DOTs and other agencies need to identify unhealthy detectors based on observed traffic data.

In the transportation literature, there are very few studies on the sensor health problem. The study by Turochy and Smith (2000) assesses a detector's health based on the time series of flow and occupancy measurements. The proposed method places thresholds on the maxima of occupancy and volume, the numbers of samples with non-zero volume but zero speed, and the average effective vehicle lengths. A sensor's health is determined by the total number of its faulty records. The study by Chen et al. (2003) developed a similar method for determining sensor health by using four statistics: the number of samples with zero occupancy, the number of samples with zero flow and non-zero occupancy, the number of samples with extremely high occupancy, and the variance of flow and occupancy. The four statistics are calculated every day per sensor. The algorithm makes decisions by comparing the statistics with the predefined thresholds. Furthermore based on a classification algorithm, PeMS categorizes a sensor's health status into ten different diagnostic states, such as "line


Figure 2.1: Configuration of California loop detector network in District 7
Source: Rajagopal and Varaiya (2007)
down", "controller down", and "high value" (Rajagopal and Varaiya, 2007).

The aforementioned methods are all designed to solve the sensor health problem for individual sensors, by examining whether the data produced by an individual sensor
looks statistically correct. Thus we refer to these methods as local methods. Such methods have two major limitations. First, the thresholds in the algorithms could be challenging to determine. In practice, the thresholds may vary by locations and are subject to exogenous factors such as traffic incidents, constructions, and weather conditions. Second, all existing local methods focus on each sensor's performance individually and fail to take into account the correlations between neighboring sensors in a road network. While admitting the effectiveness of local methods, there is still a need to develop a global method which uses network flow conservation to provide additional information from a new perspective.

### 2.2 Kinematic wave theory of traffic flow

The LWR model (Lighthill and Whitham, 1955; Richards, 1956) describes the spatialtemporal evolution of flow-rate, $q(t, x)$, and density, $k(t, x)$, at time $t$ and location $x$ by the following scalar conservation law,

$$
\begin{equation*}
\frac{\partial k}{\partial t}+\frac{\partial Q(k)}{\partial x}=0 \tag{2.1}
\end{equation*}
$$

where $q=Q(k)$ is the traffic fundamental diagram. In the LWR model, the initiation, propagation, and dissipation of traffic queues are described by the shock waves and rarefaction waves.

Newell (1993) showed that the original LWR model can be greatly simplified with its Hamilton-Jacobi form using a triangular fundamental diagram as shown in Figure 2.2;

$$
\begin{equation*}
Q(k)=\min \{V k, W(K-k)\}, \tag{2.2}
\end{equation*}
$$

where $V$ is the free-flow speed, $W$ is the shock wave speed in congested traffic, $k_{c}$ is the critical density, $C$ is the capacity, and $K$ is the jam density. The traffic state with density less than $k_{c}$ is usually defined as free-flow, density equals $k_{c}$ as critical state, density larger than $k_{c}$ as congested state.


Figure 2.2: Triangular fundamental diagram

This research introduced the cumulative flow, $N(t, x)$, as the new state variable which is the number of vehicles passing $x$ and $t$ in a spatial-temporal domain. Since the flow-rate is the time-derivative of the cumulative flow; i.e.,

$$
\begin{equation*}
q(t, x)=\frac{\partial N(t, x)}{\partial t} \tag{2.3}
\end{equation*}
$$

and the density is the negative space-derivative; i.e.,

$$
\begin{equation*}
k(t, x)=-\frac{\partial N(t, x)}{\partial x} \tag{2.4}
\end{equation*}
$$

the LWR model is equivalent to the following Hamilton-Jacobi equation:

$$
\begin{equation*}
\frac{\partial N(t, x)}{\partial t}=Q\left(-\frac{\partial N(t, x)}{\partial x}\right) . \tag{2.5}
\end{equation*}
$$

For a homogeneous road segment from $x=0$ to $x=l$, we denote the cumulative traffic counts at time $t$ at the upstream and downstream boundaries by $F(t)$ and $G(t)$, respectively. If we consider the vehicle passing the downstream boundary at $t=0$ as the reference vehicle, then

$$
\begin{align*}
G(t) & =N(t, l),  \tag{2.6}\\
F(t)+n_{0} & =N(t, 0), \tag{2.7}
\end{align*}
$$

where $F(0)=G(0)=0$, and $n_{0}$ is the initial number of vehicles on the road segment.

If initially there exists no transonic rarefaction wave on the road segment; i.e., if the initial traffic condition is one of the three types: (i) the whole road is uncongested, (ii) the whole road is congested, or (iii) an upstream part of the road is uncongested and the downstream part is congested, it is shown by Jin (2015) that the following Newell's simplified kinematic wave model can be derived from the Hopf-Lax formula for the spatial-temporal domain $x>\max \{V t, l-W t\}$,

$$
\begin{equation*}
N(t, x)=\min \left\{F\left(t-\frac{x}{V}\right)+n_{0}, G\left(t-\frac{l-x}{W}\right)+K \cdot(l-x)\right\} . \tag{2.8}
\end{equation*}
$$

The Newell's model for the U-shape spatial-temporal domain is illustrated in Figure 2.3 and the light blue area indicates the domain where Newell's model applies.

Recent studies proposed several methods to solve the Hamilton-Jacobi equation (2.5) with more general concave fundamental diagram given initial and boundary conditions in the form of cumulative flow. Daganzo 2005a b) provided analytical and numerical solution to (2.5) using variational theory. In (Evans, 2010), the Hopf-Lax formula is used to solve (2.5). A solution method based on viability theory is proposed by Claudel and Bayen (2010a b).


Figure 2.3: Newell's model in the U-shape spatial-temportal domain

### 2.3 Traffic parameter and state estimation

The traffic estimation process estimates various traffic state variables (e.g. speed, density) from available data (e.g. loop detector data, GPS data) in a road network. Ideally, an estimation method should provide a complete picture of the traffic states based on limited available data (Wang and Papageorgiou, 2005).

A review of the literature identifies a number of efforts related to traffic state estimation. One class of the models relies mainly on statistical tools to estimate the traffic states based on historical data without using traffic flow models. The common techniques used in these models include:

- time series analysis: analyze the time series of the target traffic states and model them as functions of historical observations. The classic approach is regression analysis Moorthy and Ratcliffe, 1988; Van Der Voort et al., 1996; Williams and Hoel, 2003; Cetin and Comert, 2006).
- machine learning: most of them are supervised learning algorithms. They can be trained to learn a relation between the input feature and target traffic state. The most used approaches are, for example, artificial neural networks (ANNs)(Dougherty and Cobbett, 1997; Jun and Ying, 2008) and support vector machines (SVM) (Dougherty and Cobbett, 1997; Su et al., 2007).

Another class of models are developed based on traffic flow models. Many existing estimation models adopt the Kalman filter in their framework to incorporate Lagrangian data and Eulerian data.

The Kalman filter (Kalman, 1960) is a recursive estimator which uses a system model and multiple sequential observations to form state estimations. In the case of traffic state estimation, the system model is usually the traffic flow models using Eulerian data and the observations are provided by the Lagrangian sensors. The Kalman filter can be conceptualized as two phases, "predict" and "correct". In the "predict" phase, the current prior estimate is calculated based on the system equation using the previous posterior estimate. The "correct" phase combines the current prior estimate with the current observation to generate the current posterior estimate. This process repeats the two phases at each time step. An illustration of the Kalman filter for a discrete linear system is shown in Figure 2.4 , where $\hat{x}_{k}^{-}$is the current prior estimate, $\hat{x}_{k-1}$ is the previous posterior estimate, $z_{k}$ is the current observation, $u_{k}$ is the current input, and $\hat{x}_{k}$ is the current posterior estimate. The error covariance at time step $k$, $P_{k}$, is also updated in a similar fashion to represent the reliability of the observations. The standard Kalman filter only works for linear system. A major modeling challenge of the traffic state estimation using Kalman filter is how to adapter the filter to nonlinear traffic flow models.

For the purpose of velocity field estimation, Work et al. (2008) developed an ensemble


Figure 2.4: A linear discrete Kalman filter
Source: Welch and Bishop 1995)

Kalman filter to incorporate both mobile sensor and loop detector data based on the cell transmission model (CTM) (Daganzo, 1994), which is a numerical scheme of the LWR model. This study uses the vehicle speed as the state variables with a Greenshield's fundamental diagram(Greenshields and Bibbins, 1935). The ensemble Kalman filter can be considered as a Monte Carlo approximation of the Kalman filter, which updates the estimate in time by integrating a collection of possible system state. Compared with the standard Kalman filter, the ensemble integration can capture highly nonlinearities relatively well at the cost of increased computation time.

The research by Herrera and Bayen (2008) developed a Kalman filter to combine both Lagrangian and Eulerian data. To resolve the nonlinearity caused by the CTM model, the method implemented a piece-wise linear version of the CTM, switching-mode model Muñoz et al., 2003). In this case, the standard Kalman filter techniques can be used to describe the involution of the system in each condition with the corresponding system model. In addition, the study also tested the nudging method (Newtonian
relaxation) for data fusion, which relaxes the flow conservation of the LWR model by adding source terms. While implementing the CTM based on Eulerian sensor data, the source terms are inserted to "nudge" the cell density towards the Lagrangian observations. The two methods were evaluated with real world traffic data, and it was found that the Kalman filtering slightly outperformed the nudging method at the cost of being more complicated to tune and implement.

The study by Deng et al. (2013) proposed an estimation method based on a stochastic three-detector model (STD) to incorporate heterogeneous data sources. The method can be viewed as the "correct" step of the Kalman filter, where a number of linear measurement equations are derived to map the traffic measurements as functions of the state variables (cumulative flow at both ends of the road segment). The state estimation is then formulated as an optimization problem to minimize difference between the estimated cumulative counts with observed cumulative counts at the sensor's location. The Newell's simplified kinematic wave models are used as constraints.

The second order traffic flow model, for example, Payne's model (Payne, 1971), extends the LWR model to consider the fact that the vehicles have finite acceleration and deceleration rate. Nanthawichit et al. (2003) adopted Payne's traffic flow model for traffic state estimation. A Kalman filtering estimation framework was developed to combine data from both fixed location sensors and probe sensors. However, the weighting factor is fixed to be 0.5 , which assigns equal weight to both Lagrangian and Eulerian data sources at all time. The study by Wang and Papageorgiou (2005) designed an extended Kalman-filtering method to estimate traffic state based on Eulerian sensors. The second order model, MATE(Papageorgiou et al., 1990), is used to describe the traffic dynamics. The extended Kalman filter applies Taylor expansion at each time step to linearize the traffic flow model around the current system states such that the standard linear Kalman filter equations can be used.

The study by Claudel and Bayen (2010a,b) adopted the Hamilton-Jacobi form of the LWR model and developed a traffic estimation using cumulative flows as state variables. Different from previous reviewed studies, the estimation method consider Lagrangian observations as "internal boundary condition". Using the viability theory, the traffic state (cumulative count) at arbitrary point inside the study domain can be found using a minimization principle. Compared with the numerical scheme of the LWR model, e.g. CTM, this approach guarantees an exact solution with piece-wise linear initial and boundary data.

## Chapter 3

## Network Sensor Health Problem

### 3.1 Introduction

In this chapter, we propose to solve the sensor health problem on the network level based on the traffic flow conservation. Thus we refer to this problem as the Network Sensor Health Problem (NSHP). When all sensors work flawlessly, the traffic conservation principle should be strictly followed. In practice, the cumulative flow ${ }^{11}$ over a sufficiently long period of time (e.g. daily flow) should roughly follow the conservation law unless the existence of unhealthy sensors. Therefore, it is possible to pick out unhealthy sensors based on the violation of flow conservation. In the study by Waller et al. (2008), an index of network consistency was introduced to describe the agreement in the conservation of cumulative vehicle counts between neighboring sensors at a node. However, this problem is still local for individual nodes. In this study, the proposed method evaluates the flow consistency among observed traffic flows in the whole network and assign health index for each individual sensors. Thus the NSHP is more versatile, as it applies even not all links at a node are monitored

[^0](equipped with flow sensors), and more powerful as it combines information from multiple sensors.

The network flow conservation principle leads to a system of linear equations in terms of link flows, from which one can separate links into base links and non-base links, such that flows on non-base links can be derived from those on base links. The way to separate base and non-base links is generally not unique. In this study, we aim to select base and non-base links such that the difference between observed and estimated non-base link flows is minimized. In this way, the link flows on the base links are the most consistent with respect to network flow conservation. The sensors on the corresponding non-base links are likely to be unhealthy. As a feature of this optimization problem, we usually have multiple solutions. Thus we define the health index for an individual sensor as the frequency that it appears in the solution. Unlike the binary "health label" (healthy v.s. unhealthy) used in most existing methods, the health index is a number in the range of zero to one. The health indices can provide guidelines for researchers to filter unreliable traffic data and for transportation agencies to prioritize tasks for repairing and replacing sensors.

The rest of the chapter is organized as follows. In Section 3.2, we briefly review the node-based flow conservation formulation, which was proposed by $\mathrm{Ng}(\overline{2012})$ to solve the network sensor location problem (NSLP), and formulate the NSHP as a combinatorial optimization problem. In Section 3.3, we first examine the brute-force method to find all solutions and then propose a greedy search algorithm to find a subset of solutions more effectively. In Section 3.4, we solve the NSHP for a freeway network on SR-91 and compare the results with the health statuses provided by PeMS. In Section 3.5 we summarize the results of this research.

### 3.2 Formulation

The NSHP formulation is developed based on a node-based formulation of network flow conservation, which was also used by Ng (2012) to solve the Network Sensor Location Problem (NSLP). We try to identify problematic sensors by comparing the estimated link flows with the observations. The intuition behind this is, the flow estimated from the healthy sensors measurements are expected to be more consistent with observations. The NSLP is an observability problem to determine the optimal allocation of counting sensors to estimate all link flows (i.e. to achieve full link observability) (Hu et al., 2009). It is also referred to as the link observability problem by Castillo et al. (2010). In most cases, the NSLP is identified as a sub-problem of O-D estimation (Gan et al., 2005, Gentili and Mirchandani, 2011), rather than a stand-alone problem. This may explain why the first proposed solution maintains the requirements of path enumeration ( Hu et al., 2009), which is a common requirement for O-D estimation problems. In large-scale networks, this requirement becomes impractical. The recent work of Ng (2012) proposed a node-based approach which does not require an explicit enumeration of the routes. In this section, we first define base and non-base links according to network flow conservation and then formulate the NSHP as an optimization problem to minimize the inconsistency among observed link flows. Then we define the health index for each sensor. We consider a traffic network $G=(N, L)$, where $N$ is the set of non-centroid, and $L$ is the set of links. The centroids are the nodes where traffic originates/is destined to, and non-centroids denotes all the other nodes where its in-flux equals out-flux. In general, we can acquire $G$ by removing all centroids in a general traffic network. The node-link incidence
matrix of $G$ is denoted by $\Delta$, which has the following entries:

$$
\delta_{i j}= \begin{cases}-1 & \text { if the } j \text { th link is the outgoing link of node } i  \tag{3.1}\\ 1 & \text { if the } j \text { th link is the incoming link of node } i \\ 0 & \text { otherwise. }\end{cases}
$$

### 3.2.1 The Node-based Flow Conservation

The flow conservation at non-centroids nodes can be expressed as:

$$
\begin{equation*}
\Delta F_{L}=0 \tag{3.2}
\end{equation*}
$$

where $\Delta$ is the node-link incidence matrix of network $G$ formed by non-centroid nodes, and $F_{L}$ is the vector of flows on all links. Note that the flow conservation is only valid for non-centroid nodes. Suppose that the columns in $\Delta$ can be grouped into two sub-matrices $\Delta_{K^{\prime}}$ and $\Delta_{K}$, whose columns correspond to two sets of links, $K^{\prime}$ and $K$, respectively, such that $\Delta_{K^{\prime}}$ is an invertible square matrix of dimension $|N| \times|N|$. Thus the dimension of $\Delta_{K}$ is $|N| \times(|L|-|N|)$. Then the elements in $F_{L}$ can also be grouped correspondingly and (3.2) can be rewritten as

$$
\left(\begin{array}{cc}
\Delta_{K^{\prime}} & \Delta_{K} \tag{3.3}
\end{array}\right)\binom{F_{K^{\prime}}}{F_{K}}=0
$$

The existence of such invertible sub-matrix, $\Delta_{K^{\prime}}$, is proved as Proposition 1 in $(\mathrm{Ng}$, 2012).

Since $\Delta_{K^{\prime}}$ is invertible, it immediately follows that

$$
\begin{equation*}
F_{K^{\prime}}=-\Delta_{K^{\prime}}^{-1} \Delta_{K} F_{K} . \tag{3.4}
\end{equation*}
$$

That is, if one observes $F_{K}$, then $F_{K^{\prime}}$ can be calculated as $-\Delta_{K^{\prime}}^{-1} \Delta_{K} F_{K}$. Thus we refer to $K$ as a base set of $(|L|-|N|)$ links, and $K^{\prime}$ as a non-base set of $|N|$ links; correspondingly, the links are called base and non-base links. According to $3.4, K$ is a base set if and only if $\Delta_{K^{\prime}}$ is a full rank matrix. The NSLP is solved if one finds a base set; i.e., all link flows become observable if sensors are installed on a base set of links ( $\mathrm{Ng}, 2012$ ).

Note that, however, there can be multiple base sets, $K$. We denote the collection of base sets by $\mathbb{B}$. To obtain a base set, one can use the Gaussian elimination method as in (Ng, 2012) to find the linearly independent columns which form a full rank matrix $\Delta_{K^{\prime}}$.

### 3.2.2 Sensor Health Index

Based on the network flow conservation formulation in Section 3.2, we formulate the NSHP and calculate the sensor health indices for individual sensors. Consider a trimmed traffic network $G=(N, L)$ with flow sensors installed on link set $M$, i.e. the link flows on $M, F_{M}$, are observable. We further assume a base flow set, $F_{K}$, is observable. That is,

$$
\left\{\begin{array}{l}
K \in \mathbb{B}  \tag{3.5}\\
K \subseteq M
\end{array}\right.
$$

We call $M$ and $F_{M}$ as monitored links and observed flow sets, respectively. Then $M \backslash K$ is the set of monitored non-base links. In reality, the flow counting sensors tend to cover a substantial number of links. Thus, it is highly likely that multiple base link sets are observable. The relation between the link sets are shown in Figure 3.1.

In reality, (3.4) is not true due to the existence of random of systematic errors. The


Figure 3.1: Illustration of relations between link sets
more realistic version is,

$$
\begin{equation*}
\hat{F}_{K^{\prime}}=-\Delta_{K^{\prime}}^{-1} \Delta_{K} F_{K} \neq F_{K^{\prime}} \tag{3.6}
\end{equation*}
$$

To measure the inconsistency among flows on the base links, we propose to compare the calculated non-base flows with the observed non-base flows. However, only the link flows on $M$ are observable due to the allocation of the flow counting sensors. To cope with this situation, we compare the difference in observed non-base link flows $F_{M \backslash K}$, which is obtained by removing the unobserved link flows from $F_{K^{\prime}}$. As a general measurement of inconsistency for an arbitrary base set $K$, the sum of squared errors for the observed non-base link flows is calculated as follows:

$$
\begin{equation*}
S S E\left(K \mid F_{M}\right)=\left(\hat{F}_{M \backslash K}-F_{M \backslash K}\right)^{T}\left(\hat{F}_{M \backslash K}-F_{M \backslash K}\right) . \tag{3.7}
\end{equation*}
$$

After adding the constraints, the NSHP is formulated as (3.8).

$$
\begin{gather*}
\min _{K} S S E\left(K \mid F_{M}\right)  \tag{3.8a}\\
\text { s.t. } K \subseteq M  \tag{3.8b}\\
 \tag{3.8c}\\
K \in \mathbb{B}
\end{gather*}
$$

The objective function (3.7) measures the inconsistency between calculated and observed flows on monitored, non-base links so as to identify an optimal base set,
$K$. Then, the flow counting sensors on the rest of the links, $K^{\prime}$, are considered to be suboptimal. Constraint (3.8b) states that links in $K$ must be monitored. Constraint (3.8c) requires $K$ to be a base set. From the formulation, we can see that the NSHP cannot be solved when none of the base link sets are covered by the sensors.

The NSHP is a combinatorial optimization problem, and it usually has multiple optimal solutions. Thus we can define $\mathbb{B}^{*}$ as the collection of all optimal sets,

$$
\begin{equation*}
\mathbb{B}^{*}=\left\{K^{*}: \arg \min S S E\left(K^{*} \mid F_{M}\right)\right\}, \tag{3.9}
\end{equation*}
$$

where $K^{*}$ is an optimal set which satisfies both (3.8b) and (3.8b).

The health index for a sensor is then defined as the frequency that it appears in the optimal sets, formally:

$$
\begin{equation*}
H I_{i}=\frac{\sum_{j=1}^{\left|\mathbb{B}^{*}\right|} \theta_{i j}}{\left|\mathbb{B}^{*}\right|} \tag{3.10}
\end{equation*}
$$

where $H I_{i}$ is the health index of sensor $i,\left|\mathbb{B}^{*}\right|$ is the total number of optimal sets. The indicator $\theta_{i j}$ is defined as:

$$
\theta_{i j}= \begin{cases}1 & \text { sensor } i \text { is in optimal set } K_{j}^{*}  \tag{3.11}\\ 0 & \text { sensor } i \text { is not in optimal set } K_{j}^{*}\end{cases}
$$

To demonstrate the calculation of the health index, let us consider a simple network in Figure 3.2. The links and nodes are denoted by their IDs and the traffic directions are marked by the arrows. The solid lines stand for the monitored links, $M=\{1,2,4,5,6\}$.


Figure 3.2: Sample Network

The node-link incidence matrix of this network is

Link 1 Link 2 Link 3 Link 4 Link 5 Link 6
$\Delta=\begin{aligned} & \text { Node 1 } \\ & \text { Node 2 } \\ & \\ & \text { Node 3 }\end{aligned}\left(\begin{array}{cccccc}1 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1\end{array}\right)$.

Assume observed link flow vector is

$$
F_{M}=\left[\begin{array}{c}
f_{1} \\
f_{2} \\
f_{4} \\
f_{5} \\
f_{6}
\end{array}\right]=\left[\begin{array}{c}
300 \\
200 \\
200 \\
100 \\
600
\end{array}\right]
$$

A quick check shows that the observed link flows are not consistent. For example, $\hat{f}_{6}=f_{4}+f_{5}=300 \neq f_{6}=600$.

Note that the matrix after removing the corresponding columns of link 1,2, and 4,

## Link 3 Link 5 Link 6

$$
\Delta_{K^{\prime}}=\begin{align*}
& \text { Node 1 }  \tag{3.13}\\
& \\
& \\
& \\
& \text { Node 2 3 3 }
\end{align*}\left(\begin{array}{ccc}
-1 & 0 & 0 \\
1 & -1 & 0 \\
0 & 1 & -1
\end{array}\right)
$$

is full-rank. Thus $\{1,2,4\}$ constitute a base set; i.e., $K=\{1,2,4\}$. Correspondingly, $K^{\prime}=\{3,5,6\}$, and the flow for link 3,5 , and 6 can be calculated as

$$
\hat{F}_{K^{\prime}}=\left[\begin{array}{c}
\hat{f}_{3} \\
\hat{f}_{5} \\
\hat{f}_{6}
\end{array}\right]=-\left[\begin{array}{ccc}
-1 & 0 & 0 \\
1 & -1 & 0 \\
0 & 1 & -1
\end{array}\right]^{-1}\left[\begin{array}{ccc}
1 & 1 & -1 \\
0 & 0 & 0 \\
0 & 0 & 1
\end{array}\right]\left[\begin{array}{c}
300 \\
200 \\
200
\end{array}\right]=\left[\begin{array}{c}
300 \\
300 \\
500
\end{array}\right]
$$

For the monitored, non-base link set, $M \backslash K$, we have

$$
\hat{F}_{M \backslash K}=\left[\begin{array}{l}
\hat{f}_{5} \\
\hat{f}_{6}
\end{array}\right]=\left[\begin{array}{l}
300 \\
500
\end{array}\right] .
$$

Comparing with the calculated and observed flows,

$$
F_{M \backslash K}=\left[\begin{array}{c}
f_{5} \\
f_{6}
\end{array}\right]=\left[\begin{array}{l}
100 \\
600
\end{array}\right],
$$

we can calculate the SSE as,
$S S E\left(K \mid F_{M}\right)=\left(\hat{F}_{M \backslash K}-F_{M \backslash K}\right)^{T}\left(\hat{F}_{M \backslash K}-F_{M \backslash K}\right)=(300-100)^{2}+(500-600)^{2}=50000$

In Table 3.1, we list the SSE's for 8 possible base sets and their corresponding

Table 3.1: SSE for all base sets

| Number | Monitored Base Set | Monitored Non-base Set | SSE $\left(\times 10^{4}\right)$ |
| :---: | :---: | :---: | :---: |
| 1 | $\frac{1,2,4}{2,4,5}$ | $\frac{5,6}{1,6}$ | 5 |
| 2 | $1,4,5$ | 2,6 | 13 |
| 3 | $2,5,6$ | 1,4 | 13 |
| 4 | $2,4,6$ | 1,5 | 10 |
| 5 | $1,5,6$ | 2,4 | 10 |
| 6 | $\underline{1,2,5}$ | $\underline{4,6}$ | 10 |
| 7 | $1,4,6$ | 2,5 | 5 |
| 8 |  | 10 |  |

monitored non-base sets using brute-force enumeration. From the table, we can see that base sets 1 and 7 lead to the lowest flow inconsistency levels. We have found two optimal sensor sets, $\{1,2,4\}$ and $\{1,2,5\}$, where sensor 1 and 2 appeared twice, sensor 4 and sensor 5 appeared once. The optimal sets are summarized in Figure 3.3 and the health index for each sensor are calculated accordingly. We can thus conclude that sensor 1 and 2 are working fine, sensor 4 and sensor 5 might be problematic, and sensor 6 is most likely to be broken.

| ID | NSHP |  | Health Index |
| :---: | :---: | :---: | :---: |
|  | 1 | 2 |  |
| 1 |  |  | $100 \%$ |
| 2 |  |  | $100 \%$ |
| 4 |  |  | $50 \%$ |
| 5 |  |  | $50 \%$ |
| 6 |  |  | $0 \%$ |



Figure 3.3: Calculation of sensor health index

As in this example, we can solve the NSHP and calculated the health index in a relatively simple network with a brute-force method. However, the solution space would increase exponentially as the network size grows. Therefore, there is a need to develop more efficient methods to find the optimal sets and calculate the health
indices for sensors in a large road network.

### 3.3 Solution Algorithm

The NSHP is a combinatorial optimization problem to determine which sensors to be included in the "optimal base set".

### 3.3.1 The brute-force enumeration

Let us start with the brute-force enumeration algorithm. For a road network with $|M|$ monitored links, $|L|-|M|$ unmonitored links, and $|N|$ nodes, there would be $\binom{|M|}{|L|-|N|}$ candidates to check.

## Enumeration

The candidates are any combinations of choosing $|L|-|N|$ sensors from $M$.

## Evaluation

The evaluation step uses the SSE function in (3.7) to determine the score of the current candidate. The evaluation function takes the base link set $K$ and the observed link flow $F_{M}$ as the inputs. For a given set of links $K$, we first check the rank of the matrix consisted of columns from $\Delta$ corresponding to $K^{\prime}$, denoted as $\Delta_{K^{\prime}}$. If $\Delta_{K^{\prime}}$ is a full-rank matrix, $K$ is a set of base links and the non-base link flows, $F_{K^{\prime}}$ can be estimated using (3.6). By comparing the estimated and observed non-base flows, $F_{M \backslash K}$, we calculate the SSE and use it as the score of the current candidate. If $\Delta_{K^{\prime}}$ is
not a full-rank matrix, the SSE is not calculable and we set its score to be infinity. We denote the evaluation function as $\operatorname{evl}\left(K \mid F_{M}, \Delta\right)$. Based on the previous discussion,

$$
s=\operatorname{evl}\left(K \mid F_{M}, \Delta\right)=\left\{\begin{array}{ll}
S S E\left(K \mid F_{M}\right) & \text { if } \operatorname{rank}\left(\Delta_{K^{\prime}}\right)=|N|  \tag{3.14}\\
\infty & \text { if } \operatorname{rank}\left(\Delta_{K^{\prime}}\right)<|N|
\end{array},\right.
$$

where $s$ is the score of candidate $K$.

## Algorithm

We first generate all candidates in the "enumeration" step, and then calculate the score for each candidate in the "evaluation" step. The optimal sets are the ones with the lowest score. It is a straightforward approach to find all optimal sets when the number of links and nodes are small.

### 3.3.2 The greedy search algorithm

The brute-force enumeration is not viable when we have a relatively large network. In this case, a heuristic algorithm would be helpful. In this subsection, we present a tailored greedy search algorithm for this problem. A general heuristic search algorithm in its most basic form consists of the following steps (Russell and Norvig, 1995):

1. Initialization: Generate an initial candidate to the problem.
2. Evaluation: Apply an evaluation function to the current candidate to generate a score. If the current candidate has the "best" score (the definition of "best" is usually tricky in many cases), return this candidate and quit.
3. Update: We expand the search from the current candidate by identifying the
successor candidate and return to step two.

We detail each step of the proposed algorithm in the following subsections.

## Initialization

We start the search with an arbitrary monitored base link set. This is equivalent to finding $|N|$ linearly independent columns in $\Delta$ such that the remaining $|L|-|N|$ dependent columns correspond to monitored links.

Theorem 3.3.1. If the columns in $\Delta$ corresponding to $M^{\prime}, \Delta_{M^{\prime}}$, are linearly dependent, then not all unobserved link flows can be inferred from the observed link flows

Proof. We can group the columns in $\Delta$ by the observability of link flows and rewrite 3.2 as:

$$
\begin{equation*}
\Delta_{M^{\prime}} F_{M^{\prime}}=-\Delta_{M} F_{M} \tag{3.15}
\end{equation*}
$$

where $\Delta_{M}^{\prime}$ is a matrix formed by columns corresponding to unmonitored links and $F_{M}^{\prime}$ is the vector of unobserved link flows. $\Delta_{M}$ is a matrix formed by columns corresponding to monitored links and $F_{M}^{\prime}$ is the vector of observed link flows. Assume the columns in $\Delta_{M^{\prime}}$ are linearly dependent. Let $f_{i}$ be the flow of link $i$ corresponding to a dependent column in $\Delta_{M^{\prime}}$. We can apply the standard row operations to put $\Delta_{M^{\prime}}$ in its reduced row echelon form. By the definition of the reduced row echelon form (Lay, 1997), the column corresponding to link $i$ does not contain a leading coefficient, so $f_{i}$ cannot be expressed as a linear combination of observed flows alone and thus cannot be inferred from observed link flows.

To find an initial solution, we reorganize $\Delta$ by placing columns corresponding to $M^{\prime}$
to the left and the ones corresponding to $M$ to the right to form $\Delta^{*}=\left[\Delta_{M^{\prime}} \mid \Delta_{M}\right]$. We can then use Gaussian elimination to put $\Delta^{*}$ into its reduced row echelon form. According to Theorem 3.3.1, all the column in $\Delta_{M^{\prime}}$ should be independent otherwise the unobserved flows cannot be calculated and the NSHP is not well-defined. Then all the dependent columns should correspond to the monitored links. We can use the links corresponding to independent columns as the initial solution.

The Gaussian elimination has a computational complexity of $O\left(|L \| N|^{2}\right)^{2}$. The computational cost is prohibitive for large network with abundant links, however, faster algorithms are available (Gohberg et al., 1995). We denote the initialization operation as $K=\operatorname{Ini}\left(\Delta, F_{M}\right)$, where $K$ is the initial base link set.

## Evaluation

This step is exactly the same as the "Evaluation" step in Subsection 3.3.1.

## Update

Recall that a necessary condition for $K$ to be a base link set is $|K|=|L|-|N|$. For example, given a network with 16 links, 12 nodes and 10 monitored links, the base set is to choose $16-12=4$ links from 10 monitored links. The update process generates a new candidate from the current candidate while keeping the size of the base link set unchanged. We introduce a "swap" operation, which switches one observed link in the base set with another observed link in the non-base set at a time as shown in

Figure 3.4 .

[^1]We denote the swap operation as $K_{\text {new }}=\operatorname{swap}(K, i, j)$, where $K_{\text {new }}$ is the new base set generated by swapping the i-th link in $K$ with the j-th link in $M \backslash K$ (the total number of possible swaps at each step is $(|L|-|N|) \cdot(|M|-|L|+|N|))$.


Figure 3.4: Swap operation


#### Abstract

Algorithm

Starting from a initial guess, we check all possible swaps, calculate the corresponding score using the evaluation function, and use the one with the smallest score as the "seed" for the next step. Once all possible swaps in the current step yield no smaller score than that of the current "seed", the algorithm stops. The current "seed" is the solution to NSHP. The pseudo code of the algorithm is provided in Algorithm 1 . Due to the property of the greedy search, the objective function is guaranteed to decrease monotonically. However, this algorithm does not guarantee that the final candidate is the real solution. In addition, it can find at most one of many solutions.


We apply the algorithm to solve the NSHP in the sample network shown in Figure 3.2. The node-link incident matrix in (3.12) is reorganized to get

```
Algorithm 1: A Greedy Search Algorithm
    Input:
    \(\Delta\)-Node-link incidence matrix
    \(F_{M}\)-Observed link flow
    Initialize:
    \(K=\operatorname{Ini}\left(\Delta, F_{M}\right)\)
    repeat
        \(K_{\text {seed }} \leftarrow K\)
        \(s_{\text {seed }} \leftarrow \operatorname{evl}\left(K \mid F_{M}, \Delta\right)\)
        \(s^{*} \leftarrow s_{\text {seed }}\)
        for \(i=1\) to \(|L|-|N|\) do
            for \(j=1\) to \(|M|-|L|+|N|\) do
                \(K_{\text {new }} \leftarrow \operatorname{swap}(K, i, j)\)
                \(s_{\text {new }} \leftarrow \operatorname{evl}\left(K_{\text {new }} \mid F_{M}, \Delta\right)\)
                if \(s_{\text {new }}<s^{*}\) then
                    \(s^{*}=s_{\text {new }}\)
    until \(s_{\text {seed }} \leq s^{*}\);
    Output: \(K_{\text {seed }}, s_{\text {seed }}\)
```

|  | Link 3 | Link 2 | Link 6 | Link 1 | Link 4 | Link 5 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Node 1 ( | -1 | 1 | 0 | 1 | -1 | 0 |
| $\Delta^{\prime}=$ Node 2 | 1 | 0 | 0 | 0 | 0 | -1 |
| Node 3 |  | 0 | -1 | 0 | 1 | 1 |

We use boxes to indicate non-base links and a vertical line to separate columns corresponding to unmonitored links and monitored links, respectively.

It is apparent that the first three columns in $\Delta$ have full-rank. Thus the initial base set is $\{1,4,5\}$. The initial monitored non-base set is $\{2,6\}$ and the corresponding score calculated from (3.7) is 130000 . The searching process is illustrated in Figure 3.5, where the monitored non-base links are listed since their links flows are considered to be problematic, the candidates with the lowest score in each step are indicated by
a box. The calculated score is shown after each candidate. Starting from an initial candidate of $\{2,6\},\{4,6\}$ has a lower score, and is thus used as the seed for the next step. In step two, no candidate yields smaller score than $\{4,6\}$ after exhausting the swaps, so $\{4,6\}$ is the optimal set. This means that the flows on link 4 and 6 are problematic. Note that $\{5,6\}$ is also optimal since it has the same score as $\{4,6\}$. Thus we can get the same optimal sets and same health indices for all sensors.
$\{2,6\}: 13\left\{\begin{array}{l}\{1,6\}: 13 \\ \{1,2\}: \infty \\ \{4,6\}: 5^{*} \\ \{2,4\}: 10 \\ \{5,6\}: 5 \\ \{2,5\}: 10\end{array}\left\{\begin{array}{l}\{1,6\}: 13 \\ \{1,4\}: 10 \\ \{2,6\}: 13 \\ \{2,4\}: 10 \\ \{5,6\}: 5 \\ \{4,5\}: \infty\end{array}\right.\right.$

Score $\left(\times 10^{4}\right)$

Figure 3.5: Solution procedure for the sample network

Although we are fortunate to find all optimal sets in this case study, the heuristic algorithm is not guaranteed to find all optimal sets in general. This poses potential issues when calculating the sensor health index based on limited number of optimal sets. However, if the optimal sets identified by the heuristic method is totally random, the health index should still be meaningful as the healthy sensors would appear in the majority of the optimal sets.

By using a heuristic algorithm rather than the brute-force enumeration, we trade completeness for efficiency. In this algorithm, some candidates could be re-visited at a later step. For example, $\{1,6\}$ is re-visited in the second step. To avoid inducing redundant calculations, we can cache the score of visited candidates in implementation.

### 3.4 A Real World Example

In this section, we calculate the health index for sensors on a real world network with flow counts provided by PeMS. Both brute-force enumeration and the greedy search algorithm are investigated and their results compared with PeMS health statuses respectively. The network shown in Figure 3.6 represents the network of interest, which is a stretch of eastbound State Route 91 (SR-91) in southern California. The monitored links are labeled by their IDs. We use the total daily flow reported by the PeMS system on Thursday, Feb 14th, 2013.


Figure 3.6: The SR-91 east network

The selected network contains 30 links and 16 non-centroid nodes. Among them, 18 links are monitored as shown as solid lines in Figure 3.6. Here the HOV lanes are considered separate links, except link 26 where the HOV lane is continuously accessible at this location.

So the base link set contains $(30-16)=14$ links out of 18 monitored links, which results in $\binom{18}{14}=3060$ candidates to check with the enumerations. If we apply the greedy search algorithm in the preceding section, there will be totally $(18-14) \times 14=56$ candidates to search within each step.

The greedy search process is illustrated in Figure 3.7. The candidates with infinite scores are not shown for simplicity. Similar to Figure 3.5, the IDs of the monitored


Figure 3.7: Lattice of solutions
non-base links are listed. The corresponding score is provided after the ID's. Using the method in Subsection 3.3.2, we find $\{7,14,17,21\}$ as an initial candidate to start with.

In the first step, $\{3,7,14,21\}$ yields the lowest score. It is then used as the seed for next step. In step two, $\{3,7,14,26\}$ is found to have the lowest score. In step three, since all the variations have greater or equal score than $\{3,7,14,26\}$ (marked by a star sign), the algorithm stops. The optimal base set corresponding to $\{3,7,14,26\}$ leads to the minimum flow inconsistency, and the sensors on $\{3,7,14,26\}$ are considered to be problematic. The flow on the problematic links can be inferred from the base links. It is important to notice that $\{3,7,14,26\}$ is not the unique solution to the problem. In step three, we have identified another six link sets (boxed in Figure 3.7) having the same score. We can then calculate the health indices for the sensors as in Figure 3.8. Based on the health index, sensor 1 ( $29 \%$ ), 7 ( $14 \%$ ), $14(14 \%), 22$ ( $71 \%$ ), and 26 (43\%) have the lowest health index among the sensors. In this case, we are able to match all of the problematic sensors in PeMS dataset.

| ID | Location | Lane Type | NSHP (Greedy Search) |  |  |  |  |  |  | Health Index | Pems | Pems Health Status $\{$Optimal Set $\{$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |  |  |  |  |
| 1 | GILBERT | Mainline |  |  |  |  |  |  |  | 86\% |  |  |  |
| 2 | GILBERT | HOV |  |  |  |  |  |  |  | 86\% |  |  |  |
| 3 | BROOKHURST | Off Ramp |  |  |  |  |  |  |  | 29\% |  |  |  |
| 6 | BROOKHURST | HOV |  |  |  |  |  |  |  | 86\% |  |  |  |
| 7 | BROOKHURST | Mainline |  |  |  |  |  |  |  | 14\% |  |  |  |
| 8 | BROOKHURST | On Ramp |  |  |  |  |  |  |  | 100\% |  |  |  |
| 10 | EUCLID | Off Ramp |  |  |  |  |  |  |  | 100\% |  |  |  |
| 13 | EUCLID | HOV |  |  |  |  |  |  |  | 86\% |  |  |  |
| 14 | EUCLID | Mainline |  |  |  |  |  |  |  | 14\% |  |  |  |
| 15 | EUCLID | On Ramp |  |  |  |  |  |  |  | 100\% |  |  |  |
| 17 | HARBOR 1 | Off Ramp |  |  |  |  |  |  |  | 100\% |  |  |  |
| 20 | LEMON | HOV |  |  |  |  |  |  |  | 100\% |  |  | Healthy High Value Unstable |
| 21 | LEMON | Mainline |  |  |  |  |  |  |  | 100\% |  |  |  |
| 22 | LEMON | On Ramp |  |  |  |  |  |  |  | 71\% |  |  |  |
| 24 | EAST | Off Ramp |  |  |  |  |  |  |  | 86\% |  |  |  |
| 26 | EAST | Mainline |  |  |  |  |  |  |  | 43\% |  |  |  |
| 27 | ACACIA | HOV |  |  |  |  |  |  |  | 100\% |  |  | In optimal set |
| 29 | EAST | On Ramp |  |  |  |  |  |  |  | 100\% |  |  | Not in optimal set |

Figure 3.8: Health Index (Greedy Search) v.s. PeMS

To confirm the effectiveness of the proposed method on this test network, we used brute-force enumeration and found all 12 optimal sets from the 3060 possible candidates. The results are listed in Figure 3.9. The first seven solutions are discovered by the proposed heuristic search algorithm. The health indices are calculated based on the frequency that the corresponding sensor appeared in the optimal set. The solutions of the NSHP is compared with PeMS health indices in the last column.

If we rank the sensors by their corresponding health indices, we can clearly see that all of the "high value" sensors have the lowest health indices. By using the network flow conservation information, we are able to match all "high value" sensors successfully. Although the proposed method assigned the same health index ( $67 \%$ ) for three other healthy sensors, the "unstable" sensor still has the fifth lowest health index among all sensors. The detection error regarding the "unstable" sensor might be caused by the duration of the aggregation time. Since we use accumulated daily flow in this case, the short-period fluctuations in flow counts could cancel out with each other when summed up.

| ID | Location | Lane Type | NSHP (Brute-force) |  |  |  |  |  |  |  |  |  |  |  | Health Index | Pems |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |  |  |  |  |
| 1 | GILBERT | Mainline |  |  |  |  |  |  |  |  |  |  |  |  | 67\% |  |  |  |
| 2 | GILBERT | HOV |  |  |  |  |  |  |  |  |  |  |  |  | 83\% |  |  |  |
| 3 | BROOKHURST | Off Ramp |  |  |  |  |  |  |  |  |  |  |  |  | 50\% |  |  |  |
| 6 | BROOKHURST | HOV |  |  |  |  |  |  |  |  |  |  |  |  | 67\% |  |  |  |
| 7 | BROOKHURST | Mainline |  |  |  |  |  |  |  |  |  |  |  |  | 33\% |  |  |  |
| 8 | BROOKHURST | On Ramp |  |  |  |  |  |  |  |  |  |  |  |  | 100\% |  |  |  |
| 10 | EUCLID | Off Ramp |  |  |  |  |  |  |  |  |  |  |  |  | 100\% |  |  |  |
| 13 | EUCLID | HOV |  |  |  |  |  |  |  |  |  |  |  |  | 67\% |  |  |  |
| 14 | EUCLID | Mainline |  |  |  |  |  |  |  |  |  |  |  |  | 33\% |  |  |  |
| 15 | EUCLID | On Ramp |  |  |  |  |  |  |  |  |  |  |  |  | 100\% |  |  |  |
| 17 | HARBOR 1 | Off Ramp |  |  |  |  |  |  |  |  |  |  |  |  | 100\% |  |  | Healthy |
| 20 | LEMON | HOV |  |  |  |  |  |  |  |  |  |  |  |  | 100\% |  | Pems Health status | High Value |
| 21 | LEMON | Mainline |  |  |  |  |  |  |  |  |  |  |  |  | 100\% |  |  | Unstable |
| 22 | LEMON | On Ramp |  |  |  |  |  |  |  |  |  |  |  |  | 67\% |  |  |  |
| 24 | EAST | Off Ramp |  |  |  |  |  |  |  |  |  |  |  |  | 92\% |  |  |  |
| 26 | EAST | Mainline |  |  |  |  |  |  |  |  |  |  |  |  | 42\% |  |  |  |
| 27 | ACACIA | HOV |  |  |  |  |  |  |  |  |  |  |  |  | 100\% |  | SHP Health Index | Healthy |
| 29 | EAST | On Ramp |  |  |  |  |  |  |  |  |  |  |  |  | 100\% |  | NSHP Health index | Unhealthy |

Figure 3.9: Health Index (Brute-force) v.s. PeMS

### 3.5 Summary

In this chapter, we formulated the network sensor health problem (NSHP) to find the optimal base set of sensors whose measurements are the most consistent with respect to the network flow conservation. We proposed the health index as a measurement to indicate individual sensor's health condition based on identified optimal sets. While the brute-force enumeration can be used to find all optimal sets, we developed a tailored greedy search algorithm to find some of the optimal sets effectively. The formulation is further tested on a real world network to compare with the sensor health status from PeMS. We found a perfect match between the health indices and the results provided by PeMS. This indicates that the network flow conservation information is extremely useful to evaluate sensor health and equally important as any local information.

## Chapter 4

## Traffic Estimation with Vehicle Reidentification Data

### 4.1 Introduction

Despite the substantial progress made in traffic estimation, existing estimation methods are limited since model parameters and initial states have to be predetermined. First, most of the traffic flow models used by existing methods involve a number of model parameters, especially those in the fundamental diagram, including the free-flow speed, jam density, and shock wave speed. In most existing studies, the fundamental diagram is assumed to be given, or calibrated before traffic estimation. Second, in some of the existing studies, the initial states are assumed to be either empty (Deng et al., 2013) or known (Sun et al., 2003). Such assumptions can limit the accuracy of the results as well as their transferability to different locations and times.

In this chapter, we attempt to fill the gap by proposing a new method to simultaneously estimate model parameters and traffic states on a homogeneous link with loop detector
data and vehicle reidentification data. We refer to it as the LoopReid method thereafter. Note that, by simultaneous estimation, we mean that model parameters in the fundamental diagram, the initial states, and later traffic state can all be estimated within the same framework, but calculations can still be sequentially ordered. This is substantially different from existing methods where model parameters and initial states have to be observed or estimated with other methods. The simultaneous estimation of both parameters and states is achieved by formulating a single optimization problem in terms of the initial number of vehicles and model parameters. Here we assume that both Eulerian traffic count data and a portion of Lagrangian vehicle trajectory data are available through loop detectors, and vehicle reidentification systems or GPS devices, respectively. We then decouple the optimization problem such that the initial number of vehicles can be calculated with a closed-form formula, and the model parameters, including the jam density and shock wave speed in congested traffic, can be computed with the Gauss-Newton method. Further, based on Newell's model, we can calculate the average densities of the segment as well as speeds, and flow-rates inside the road segment.

The proposed framework has some similarities with the stochastic three detector (STD) method proposed in (Deng et al. 2013): both are based on Newell's simplified kinematic wave model and use the same types of data, including Eulerian traffic counts and Lagrangian vehicle trajectories. However, the two methods are fundamentally different. Deng et al. (2013) assumes predetermined model parameters and initial states, and the resulting optimization problems are substantially different.

The rest of the chapter is organized as follows. In Section 4.2, we present the new estimation framework based on Newell's model for a congested road segment and formulate an optimization problem in the initial traffic state and model parameters. In Section 4.3, we present the solution method and discuss some properties of the
optimization problem. In Section 4.4, we use the NGSIM data (USDOT, 2008) to test the validity of our method with different correct matching rates. As a benchmark, we compare our method with the STD method which uses the same dataset. We conclude the this chapter in Section 4.5 .

### 4.2 The LoopReid estimation method



Figure 4.1: Illustration of a vehicle trajectory and kinematic waves on a road segment

The proposed traffic estimation method is developed based on Newell's simplified kinematic wave model (2.8) using Loop detector data and vehicle reidentificaiton data. The Eulerian traffic count data, $F(t)$ and $G(t)$, are available through loop detectors at the two ends of a road segment; in addition, Lagrangian vehicle data, the entry and exit times of vehicles, are available through vehicle reidentification technologies. We assume that $I$ vehicles are reidentified. If vehicle $i(i=1, \cdots, I)$ is reidentified at the downstream boundary, we denote its entry and exit times in a road segment by $r(i)$ and $s(i)$, respectively. If we denote $X(t, i)$ as the location of vehicle $i$ at $t$,
then $X(r(i), i)=0$ and $X(s(i), i)=l$. Here we assume only a portion of vehicles are reidentified. In Figure 4.1, we illustrate the trajectory of vehicle $i$ (the solid curve), whose entry and exit times are known, and the kinematic waves in uncongested and congested traffic respectively (the dashed lines).

### 4.2.1 An optimization problem in initial states and parameters

Assume that vehicles follow the First-In-First-Out (FIFO) principle on the whole road segment, then the contour lines of $N(t, x)$ are vehicle trajectories. In particular, the cumulative numbers of vehicles should be equal when a reidentified vehicle passes the upstream and downstream boundaries; i.e.,

$$
\begin{equation*}
F(r(i))+n_{0}=G(s(i)) . \tag{4.1}
\end{equation*}
$$

From (2.7) and (2.8), we have

$$
\begin{align*}
F(r(i))+n_{0} & =N(r(i), 0)=\min \left\{F(r(i))+n_{0}, G\left(r(i)-\frac{l}{W}\right)+K l\right\}  \tag{4.2}\\
G(s(i)) & =\min \left\{F\left(s(i)-\frac{l}{V}\right)+n_{0}, G(s(i))\right\} \tag{4.3}
\end{align*}
$$

Combining the above equations with (4.1) we further have

$$
\begin{align*}
G(s(i)) & =\min \left\{F(r(i))+n_{0}, G\left(r(i)-\frac{l}{W}\right)+K l\right\},  \tag{4.4}\\
F(r(i))+n_{0} & =\min \left\{F\left(s(i)-\frac{l}{V}\right)+n_{0}, G(s(i))\right\} . \tag{4.5}
\end{align*}
$$

In particular, when traffic is congested, traffic information propagates upstream with a speed of $-W$, and we have

$$
\begin{equation*}
G(s(i))=G\left(r(i)-\frac{l}{W}\right)+K l \tag{4.6}
\end{equation*}
$$

when traffic is uncongested, traffic information propagates downstream with a speed of $V$, and we have

$$
\begin{equation*}
F(r(i))+n_{0}=F\left(s(i)-\frac{l}{V}\right)+n_{0} . \tag{4.7}
\end{equation*}
$$

In this study, we focus on congested traffic conditions, as the uncongested traffic condition is less of a concern in practice. Therefore, we will use (4.6) for further discussion, but the results for (4.7) will be similar.

Note that the equalities in (4.1) and (4.6) can be altered by the following four types of errors:

1. Errors caused by the measurement of cumulative traffic counts. Cumulative traffic counts are usually collected with loop detectors and can be corrupted by noise and errors, due to a variety of problems including pavement/saw-cut failures, intermittent communications, double counting of lane-changing vehicles, and so on (Coifman, 2006; Lee and Coifman, 2011). According to PeMS, which is a widely used data source for the freeway sensor system in California, only $67 \%$ of the detectors were working properly in May, 2014. Some districts (e.g. district 7 in Los Angeles County) have even lower proportions of working detectors.
2. Errors caused by the vehicle reidentification technology. The entry and exit times of reidentified vehicles can be corrupted by time shifts and vehicle mismatching in the vehicle reidentification technology. The time shift is mainly caused by the spatial and temporal inaccuracy of the underlying detection technology, including the Bluetooth technology or the GPS technology (Wang et al., 2011).

Mismatch errors can be caused by a variety of algorithmic and technology limitations. Jeng (2007) defined the correct matching rate (CMR) as the ratio between total number of correct matched vehicles and total number of vehicles, and obtained CMR of $69.76 \%$ to $73.59 \%$ in off-peak conditions and $50.68 \%$ to $54.20 \%$ in peak conditions.
3. Errors caused by the violation of the FIFO principle. The FIFO principle is a fundamental assumption to derive both (4.1) and (4.6). In reality, however, the principle is only approximately followed on a multi-lane road, as travel speeds of different types of vehicles on different lanes are usually different in the same longitudinal location at the same time. In the research by Jin and Li (2007), it was confirmed that the FIFO principle is generally violated, and the level of FIFO violation was empirically studied with an NGSIM I80 database.
4. Errors associated with Newell's simplified kinematic wave theory. Newell's simplified kinematic wave theory, (2.8), is valid only under a number of assumptions: (i) there exists an equilibrium relation between flow-rate and density; (ii) the fundamental diagram is triangular; and (iii) initially there exists no transonic rarefaction wave. In reality, however, (i) the existence of an equilibrium flow-density relation is only approximately true, as the observed data are be quite scattered (Hall et al., 1986); (ii) the shape of the fundamental diagram is also debatable, due to impacts of lane changes, multi-class vehicles, and so on (Jin, 2013); and (iii) incidents could create some transonic rarefaction waves on a road.

To carefully study the properties of the aforementioned errors can be fundamentally important, but is beyond the scope of this study. Here we introduce error terms into
both (4.1) and (4.6), such that

$$
\begin{align*}
n_{0} & =G(s(i))-F(r(i))+\epsilon_{i},  \tag{4.8}\\
G(s(i)) & =G\left(r(i)-\frac{l}{W}\right)+K l+\xi_{i} . \tag{4.9}
\end{align*}
$$

In particular, $\epsilon_{i}$ is related to the first three types of errors, and $\xi_{i}$ all four types of errors. We propose to estimate $n_{0}, W$, and $K$ by minimizing the sum of squared errors in (4.8) and (4.9) with a least square estimation framework (Charnes et al., 1976):

$$
\begin{equation*}
\min _{\hat{n}_{0}, \hat{W}, \hat{K}} \sum_{i=1}^{I} \epsilon_{i}^{2}+\xi_{i}^{2}=\sum_{i=1}^{I}\left[G(s(i))-F(r(i))-\hat{n}_{0}\right]^{2}+\left[G(s(i))-G\left(r(i)-\frac{l}{\hat{W}}\right)-\hat{K} l\right]^{2}, \tag{4.10}
\end{equation*}
$$

where $\hat{n}_{0}$ is the estimated number of vehicles initially on the road segment, and $\hat{W}$ and $\hat{K}$ are respectively estimated shock wave speed and jam density.

### 4.2.2 Estimation of traffic state

Once the initial number of vehicles and the three model parameters are estimated by solving the optimization problem, 4.10, we can estimate the traffic state in the spatio-temporal domain where the Newell's model applies $(x>\max \{V t, l-W t\})$ :

- The cumulative flows can be calculated using Newell's simplified kinematic wave model, 2.8, as follows:

$$
\begin{equation*}
\hat{N}(t, x)=\min \left\{F\left(t-\frac{x}{\hat{V}}\right)+\hat{n}_{0}, G\left(t-\frac{l-x}{\hat{W}}\right)+\hat{K} \cdot(l-x)\right\}, \tag{4.11}
\end{equation*}
$$

where $\hat{V}$ is the estimated free-flow speed which can be estimated in uncongested
traffic by solving a similar optimization problem as in 4.10.

- From the estimated cumulative flows, we can also estimate the average density inside a subsegment from $x_{1}$ to $x_{2}$ at time $t$, denoted by $\hat{k}\left(t ; x_{1}, x_{2}\right)$ :

$$
\begin{equation*}
\hat{k}\left(t ; x_{1}, x_{2}\right)=\frac{1}{x_{2}-x_{1}}\left[G\left(t-\frac{l-x_{1}}{\hat{W}}\right)-G\left(t-\frac{l-x_{2}}{\hat{W}}\right)+\hat{K}\left(x_{2}-x_{1}\right)\right] . \tag{4.12}
\end{equation*}
$$

In particular, for the whole road segment from 0 to $l$, the average density can be estimated by

$$
\begin{equation*}
\hat{k}(t ; 0, l)=\frac{1}{l}\left[F(t)-G(t)+\hat{n}_{0}\right] \tag{4.13}
\end{equation*}
$$

Note that the estimated average density for the whole road segment only involves the estimated initial number of vehicles, and does not directly depend on $\hat{K}$ or $\hat{W}$.

- We can also estimate the flow-rate at any location $x \in[0, l]$ and time $t$. Since we only consider congested traffic, the flow-rate at location $x$ is determined by the cumulative flow at the downstream:

$$
\begin{equation*}
\hat{q}(t, x)=\frac{N(t+\Delta t, x)-N(t)}{\Delta t}=\frac{G\left(t+\Delta t-\frac{l-x}{\hat{W}}\right)-G\left(t-\frac{l-x}{\hat{W}}\right)}{\Delta t} \tag{4.14}
\end{equation*}
$$

where $\Delta t$ is the time-step size. In particular, the upstream flow-rate, $f(t)=$ $q(t, 0)$, can be estimated as

$$
\begin{equation*}
\hat{f}(t)=\frac{G\left(t+\Delta t-\frac{l}{\hat{W}}\right)-G\left(t-\frac{l}{\hat{W}}\right)}{\Delta t} \tag{4.15}
\end{equation*}
$$

Alternatively, $f(t)$ can be calculated from observed cumulative traffic counts at the upstream boundary

$$
\begin{equation*}
f(t)=\frac{F(t+\Delta t)-F(t)}{\Delta t} \tag{4.16}
\end{equation*}
$$

By comparing $\hat{f}(t)$ and $f(t)$ we can measure the accuracy of the new estimation method.

In addition to the aforementioned variables, we can also estimate travel speeds, travel times, and other traffic information from the estimated cumulative flows, $\hat{N}(t, x)$.

### 4.3 Solution of the optimization problem

In this section, we present a solution method to 4.10 and discuss some of its properties.

### 4.3.1 A decoupling method

The optimization problem, 4.10), can be decoupled as follows.

Theorem 4.3.1. The optimization problem, 4.19), can be decoupled into the following two optimization problems:

$$
\begin{align*}
& \min _{\hat{n}_{0}} \sum_{i=1}^{I}\left[G(s(i))-F(r(i))-\hat{n}_{0}\right]^{2}  \tag{4.17a}\\
& \min _{\hat{W}, \hat{K}} \sum_{i=1}^{I}\left[G(s(i))-G\left(r(i)-\frac{l}{\hat{W}}\right)-\hat{K} l\right]^{2} . \tag{4.17b}
\end{align*}
$$

Proof. This is straightforward, since, in the objective function of (4.10), $\hat{n}_{0}$ appears only in the first term, and $\hat{W}$ and $\hat{K}$ appear only in the second term. Thus the objective function of 4.10) is minimized if and only if both terms are minimized.

Therefore we can solve (4.17a and 4.17b separately. In particular, since the objective function in 4.17a is quadratic in $\hat{n}_{0}$, it has the following closed-form
analytical solution.

Corollary 4.3.2. The initial number of vehicles on the road segment can be estimated by

$$
\begin{equation*}
\hat{n}_{0}=\frac{\sum_{i=1}^{I}[G(s(i))-F(r(i))]}{I} . \tag{4.18}
\end{equation*}
$$

That is, the initial number of vehicles equals the mean value of the differences in the cumulative counts at the entry and exit times of reidentified vehicles.

Proof. Taking the derivative of the objective function in 4.17a with respect to $\hat{n}_{0}$ and setting it to zero, we can easily obtain (4.18). Furthermore we can verify that the objective function attains its minimum, since the second-order derivative is positive at the optimal point.

Although the objective function in 4.17 b is quadratic in $\hat{K}$, its relation with $\hat{W}$ is not in a simple functional form. Thus (4.17b) cannot be analytically solved. Since the objective function in 4.17 b is quadratic in $\hat{K}$, with known $\hat{W}$ the optimal value of $\hat{K}$ can be analytically solved as

$$
\begin{equation*}
\hat{K}=\sum_{i=1}^{I} \frac{G(s(i))-G\left(r(i)-\frac{l}{\hat{W}}\right)}{I l} . \tag{4.19}
\end{equation*}
$$

In addition, if the flow-rate is relatively constant over time; i.e., when traffic is in a steady state (Cassidy, 1998) with an average flow-rate of $\bar{q}$, we have

$$
\begin{equation*}
G(s(i))-G\left(r(i)-\frac{l}{\hat{W}}\right) \approx\left(s(i)-r(i)+\frac{l}{\hat{W}}\right) \bar{q} \tag{4.20}
\end{equation*}
$$

In this case, the objective function in 4.17 b is also convex in $\hat{W}$. This property justifies the usage of a gradient-based Gauss-Newton method (Milliken, 1990).

We denote $\boldsymbol{\theta}$ as the column vector of $(\hat{W}, \hat{K})^{T}$. Starting from an initial guess $\boldsymbol{\theta}^{(0)}=$
$\left(\hat{W}^{(0)}, \hat{K}^{(0)}\right)^{T}$, the Gauss-Newton method updates the results by iterating

$$
\boldsymbol{\theta}^{(j+1)}=\boldsymbol{\theta}^{(j)}-\left[\boldsymbol{J}\left(\boldsymbol{\theta}^{(j)}\right)^{T} \boldsymbol{J}\left(\boldsymbol{\theta}^{(j)}\right)\right]^{-1} \boldsymbol{J}\left(\boldsymbol{\theta}^{(j)}\right)^{T} \boldsymbol{\xi}\left(\boldsymbol{\theta}^{(j)}\right),
$$

is the Jacobian matrix of $\boldsymbol{\xi}(\boldsymbol{\theta})=\left(\xi_{1}(\boldsymbol{\theta}), \xi_{2}(\boldsymbol{\theta}), \ldots, \xi_{I}(\boldsymbol{\theta})\right)^{T}$ and is defined as

$$
\boldsymbol{J}(\boldsymbol{\theta})=\left[\begin{array}{cc}
\frac{\partial \xi_{1}(\boldsymbol{\theta})}{\partial \hat{W}} & \frac{\partial \xi_{1}(\boldsymbol{\theta})}{\partial \bar{K}}  \tag{4.21}\\
\vdots & \vdots \\
\frac{\partial \xi_{I}(\boldsymbol{\theta})}{\partial \hat{W}} & \frac{\partial \xi_{I}(\boldsymbol{\theta})}{\partial \hat{K}}
\end{array}\right] .
$$

The essential idea is to approximate the Hessian matrix of the objective function, $\boldsymbol{\xi}^{T}\left(\boldsymbol{\theta}^{(j)}\right) \boldsymbol{\xi}\left(\boldsymbol{\theta}^{(j)}\right)$, with its first order approximation $\boldsymbol{J}\left(\boldsymbol{\theta}^{(j)}\right)^{T} \boldsymbol{J}\left(\boldsymbol{\theta}^{(j)}\right)$ at time-step $j$. The iteration stops when the solution converges. We consider that the algorithm has converged when the second norm of the error $\left\|\boldsymbol{\theta}^{(j+1)}-\boldsymbol{\theta}^{(j)}\right\|$ is smaller than a predefined error bound (we used $10^{-4}$ as it is a typical setting for engineering practice).

In this problem, the elements in $\boldsymbol{J}(\boldsymbol{\theta})$ are $(i=1, \cdots, I)$

$$
\begin{equation*}
\frac{\partial \xi_{i}(\boldsymbol{\theta})}{\partial \hat{W}}=-\left.\frac{l}{\hat{W}^{2}} \frac{d G(t)}{d t}\right|_{t=r(i)-\frac{l}{\hat{W}}}=-\frac{l}{\hat{W}^{2}} g\left(r(i)-\frac{l}{\hat{W}}\right), \frac{\partial \xi_{i}(\boldsymbol{\theta})}{\partial \hat{K}}=-l . \tag{4.22}
\end{equation*}
$$

The downstream flow-rate in $4.22, g\left(r(i)-\frac{l}{\hat{W}}\right)$, is approximated by the the forward finite difference,

$$
\begin{equation*}
g\left(r(i)-\frac{l}{\hat{W}}\right) \approx \frac{G\left(r(i)-\frac{l}{\hat{W}}+\Delta t\right)-G\left(r(i)-\frac{l}{\hat{W}}\right)}{\Delta t} \tag{4.23}
\end{equation*}
$$

where $\Delta t$ is the time-step size. For simplicity, we fix the time-step size to be 30 seconds.

### 4.3.2 Potential issue with near-stationary traffic

When traffic is absolutely stationary such that the trajectories of vehicles are parallel lines, we denote the density, speed, and flow-rate by $k, v$, and $q$, respectively. Here $q=k v$. Assuming that there are no measurement errors in the cumulative traffic counts, then

$$
G(s(i))-G\left(r(i)-\frac{l}{\hat{W}}\right)=\left(s(i)-r(i)+\frac{l}{\hat{W}}\right) q,
$$

Further, assuming that there are no errors caused by reidentification technologies, the travel time for reindentified vehicle $i$ is constant

$$
\begin{equation*}
s(i)-r(i)=\frac{l}{v} . \tag{4.24}
\end{equation*}
$$

Thus the optimization problem, 4.17b, can be simplified as

$$
\begin{equation*}
\min _{\hat{W}, \hat{K}} l^{2} \cdot \sum_{i=1}^{I}\left(k+\frac{q}{\hat{W}}-\hat{K}\right)^{2}, \tag{4.25}
\end{equation*}
$$

which is analytically solved by

$$
\begin{equation*}
\hat{K}=k+\frac{q}{\hat{W}} . \tag{4.26}
\end{equation*}
$$

In this case, $\hat{W}$ and $\hat{K}$ follow a hyperbolic relationship, and cannot be determined uniquely. That is, if vehicles have parallel trajectories in stationary traffic, there are multiple (infinite) solutions for $\hat{W}$ and $\hat{K}$.

In reality, it is impossible for traffic to be absolutely stationary. However, it is possible that the traffic is in near-stationary state. In this case, the objective function in 4.17b) becomes relatively flat and its value is almost constant for a large range of values in $\hat{W}$ and $\hat{K}$. Thus, the parameters are hardly identifiable. This insight was also used in calibrating car-following models (Yang et al., 2011). Therefore, such limitations are actually caused by a lack of variation in the observed data, not the
estimation framework itself.

### 4.4 Applications to NGSIM I101 dataset

In this section, the performance of the proposed method is evaluated in terms of its ability to estimate the initial traffic state $\left(n_{0}\right)$, the average density for the whole road segment, and the upstream flow-rate. Here the data are chosen from the Next Generation Simulation (NGSIM) database (USDOT, 2008) for the US 101 freeway in Los Angeles, CA, from 7:50 AM to 8:35 AM on June 15, 2005. The study site is a segment between the Ventura Blvd and the Cahuenga Blvd off-ramps on the southbound US 101 freeway as shown in Figure 4.2. The segment is about 0.13 miles long with five lanes and one auxiliary lane. Using the the interval 7:50 AM to 8:05 AM as an example, the first three minutes are used as the warm-up period. After 7:53 AM, all the vehicle trajectories are used in estimation

The 45 -min period is split into three 15 -minute intervals. In this study, we only consider the through traffic which passed both upstream and downstream of the road segment. There are 1894 vehicle trajectories in the first time period, 1842 in the second time period, and 1698 in the third period. We deploy virtual detectors at both ends of the road segment and generated volume counts and vehicle entry/exit times based on vehicles' longitudinal coordinates ("LocalY" field in the NGSIM trajectory dataset). While there has been some discussion about the NGSIM data accuracy in recent years (Choi, 2014; Punzo et al., 2011; Thiemann et al., 2008), most of the discovered issues are related to speed and acceleration. The longitudinal coordinates are unbiased and it is acceptable to use it as the actual vehicle path(Punzo et al., 2011).


Figure 4.2: Illustration of the Study Site

### 4.4.1 Data Preparation

On average, it took around 25 seconds for vehicles to go through this segment under congested conditions. Only the left five lanes of data were used since the auxiliary lane had much lower flow-rates and was significantly under-utilized, which violates the congested traffic state assumption. To study the impact from sample size of vehicle reidentification, we used four different CMRs ( $100 \%, 50 \%, 20 \%, 5 \%$ ) for each time interval; i.e., we assume respectively $100 \%, 50 \%, 20 \%$, and $5 \%$ of all vehicles are correctly reidentified at both upstream and downstream boundaries and their entry/exit times are used in the estimation. We use a 3-minute warm-up period such that all vehicle trajectories are captured in the spatial-temporal domain (see Figure $4.2)$.

The sampling frequency of NGSIM is 10 Hz . For vehicle $i$, its location at time step $j$ is $X(j \Delta t, i)$, where $\Delta t=0.1 \mathrm{~s}$. We prepared two datasets for the estimation method, namely, the vehicle reidentification measurements and the flow measurements.


Figure 4.3: Piecewise linear interpolation of cumulative flows

- The vehicle reidentification data contain the entry/exit times of individual vehicles at the study site. Let $x_{0}$ and $x_{l}$ denote the entry and exit point of the segment respectively. We used a linear interpolation function to find the entry time $s(i)$, formally,

$$
s(i)=j \Delta t+\frac{x_{0}-X(j \Delta t, i)}{X(j \Delta t+\Delta t, i)-X(j \Delta t, i)} \Delta t
$$

where $j$ satisfies $X(j \Delta t, i) \leq x_{0} \leq X(j \Delta t+\Delta t, i)$. Similarly, the exit time is

$$
r(i)=j^{\prime} \Delta t+\frac{x_{l}-X\left(j^{\prime} \Delta t, i\right)}{X\left(j^{\prime} \Delta t+\Delta t, i\right)-X\left(j^{\prime} \Delta t, i\right)} \Delta t
$$

where $j^{\prime}$ is chosen such that $X\left(j^{\prime} \Delta t, i\right) \leq x_{l} \leq X\left(j^{\prime} \Delta t+\Delta t, i\right)$.

- We can interpolate the cumulative flow based on the entry and exit times, $s(i)$ and $r(i)$, for vehicle $i$. The cumulative flow functions $F(t)$ and $G(t)$ are
practically step functions. However, it is preferable to approximate them with continuous functions which are easier to evaluate. In particular, we can find the instantaneous upstream flow-rate at time $t$ by evaluating $f(t)=$ $\lim _{\Delta t \rightarrow 0} \frac{F(t+\Delta t)-F(t)}{\Delta t}$ at almost all points, and the same for downstream flow-rate $g(t)$. Here, we use the approximation method in (Daganzo, 1997). The method approximates the step function with a piecewise linear curve passing through the crests as illustrated in Figure 4.3. In the plot, the cumulative flow function at the upstream, $F(t)$, is approximated by $\tilde{F}(t)$. The resulting function is differentiable almost everywhere except at the transition points.


### 4.4.2 Estimation of initial states and model parameters

We first estimate the initial number of vehicles on the road segment, $\hat{n}_{0}$, and the model parameters, $\hat{W}$ and $\hat{K}$, with methods presented in Section 4.3.1. In the Gauss-Newton method, the initial guess of $\hat{W}$ is 20 miles per hour and that of $\hat{K}$ is 200 vehicles per mile per lane.

The results in Table 4.1 provides mean values of $\hat{n}_{0}, \hat{W}$ and $\hat{K}$ followed by their standard deviation for 100 runs each. The true values of $n_{0}$ are shown in the parentheses in the header line. First, there is no clear correlation between the mean values of the estimation and the CMRs. The mean values are relatively constant with different CMRs. Second, the variance of the estimation increases as the CMR decreases. This trend indicates that the estimation precision is lower with fewer reidentification observations. Compared with the true values, the estimation of the initial states is very accurate; the differences in mean are less than $2 \%$ with different CMRs. The estimated shock wave speed is around 22 mph for all time periods, which is higher than that calibrated from loop detector data in the PeMS (around 15 mph ).

Table 4.1: Verification results

| $7: 50$ AM - 8:05 AM |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Case No. | CMR(\%) | $\hat{n}_{0}$ (True: 39) | $\hat{W}(\mathrm{mph})$ | $\hat{K}(\mathrm{vpm})$ |  |
| 1 | 100.00 | $38.31(0.00)$ | $20.00(0.00)$ | $156.51(0.00)$ |  |
| 2 | 50.00 | $38.33(0.23)$ | $23.51(2.37)$ | $144.67(7.10)$ |  |
| 3 | 20.00 | $38.25(0.49)$ | $23.19(3.29)$ | $145.99(9.08)$ |  |
| 4 | 5.00 | $38.32(1.02)$ | $23.39(5.11)$ | $146.78(12.98)$ |  |
| $8: 05$ AM - 8:20 AM |  |  |  |  |  |
| Case No. | CMR(\%) | $\hat{n}_{0}$ (True: 39) | $\hat{W}(\mathrm{mph})$ | $\hat{K}(\mathrm{vpm})$ |  |
| 5 | 100.00 | $38.92(0.00)$ | $25.67(0.00)$ | $141.46(0.00)$ |  |
| 6 | 50.00 | $38.91(0.25)$ | $21.81(1.71)$ | $152.85(5.65)$ |  |
| 7 | 20.00 | $39.05(0.53)$ | $21.71(2.01)$ | $153.52(6.41)$ |  |
| 8 | 5.00 | $38.98(1.18)$ | $22.46(3.05)$ | $151.60(9.21)$ |  |
| $8: 20$ AM $-8: 35 \mathrm{AM}$ |  |  |  |  |  |
| Case No. | CMR(\%) | $\hat{n}_{0}($ True: 51$)$ | $\hat{W}(\mathrm{mph})$ | $\hat{K}(\mathrm{vpm})$ |  |
| 9 | 100.00 | $51.32(0.00)$ | $21.15(0.00)$ | $157.98(0.00)$ |  |
| 10 | 50.00 | $51.34(0.33)$ | $20.90(0.69)$ | $158.97(2.36)$ |  |
| 11 | 20.00 | $51.36(0.62)$ | $20.98(1.14)$ | $158.80(3.80)$ |  |
| 12 | 5.00 | $51.32(1.34)$ | $21.45(2.15)$ | $157.52(6.97)$ |  |

At the same time, the estimated jam density is lower than that from the PeMS (around 150 vpm vs 200 vpm$)$. This discrepancy may be due to the fact that the parameters estimated here are for a road segment, while PeMS results are for individual detectors. We also notice that the variances of the estimated parameters are smaller for the later time periods, indicating higher precisions. One explanation is: the estimation method is derived under the assumption of congested traffic state, so the modeling error should be relatively low for congested traffic data. Since the traffic in this dataset is getting increasingly congested, the estimation of the later time periods are more accurate.

In Figure 4.4, we show the heat map of the objective function in 4.17b in case 5. We pick 900 points from the parameter space of $\hat{W} \in[10,40]$ and $\hat{K} \in[50,300]$ with 30 grids along each axis. Starting from the initial value $\left(\hat{W}^{(0)}, \hat{K}^{(0)}\right)=(20,200)$, the solution is found to be $(25.67,141.46)$ after 20 iterations with the Gauss-Newton method. From the figure we can see that the contour lines are roughly hyperbolic as in (4.26). This suggests that traffic is close to stationary, but still features enough


Figure 4.4: Plot of the objective function
oscillations so that the optimal point is identifiable.

### 4.4.3 Validation with respect to the average density and upstream flow-rate

To examine the performance of the LoopReid method and determine the reliability of the estimated values, the average density within the whole segment and the estimated upstream flow-rate are compared with their true values respectively in this subsection. The average density within the segment is estimated as in 4.13). The estimation equation for upstream flow-rate during time interval $\Delta t$ is shown in equation (4.15). To ensure a sufficient number of counts during the aggregation interval, $\Delta t$ is set to be 30 seconds.

We use the mean absolute percentage error (MAPE) as the measure of accuracy. The MAPE for the average density, $k$, is calculated as:

$$
\begin{equation*}
M A P E_{k}=\sum_{j=1}^{J}\left|\frac{k(j \Delta t)-\hat{k}(j \Delta t)}{k(j \Delta t)}\right|, \tag{4.27}
\end{equation*}
$$

where $J$ is the number of time-steps. The MAPE for the upstream flow-rate is calculated as:

$$
M A P E_{f}=\sum_{t=1}^{J}\left|\frac{f(j \Delta t)-\hat{f}(j \Delta t)}{f(j \Delta t)}\right|
$$

In Figure 4.5 we demonstrate the estimated and observed upstream flow-rates with different CMRs $(100 \%, 50 \%, 20 \%, 5 \%)$ for the three time intervals compared with the true values. From the plots we can see that the proposed estimation method is able to capture the general trends in the upstream flow-rates. Therefore, the estimated model parameters are reliable. Note that, for the case with CMR lower than $100 \%$, the line is just the result of one run. To better study the estimation accuracy, we conducted repeated sampling for all cases and the results are shown in Table 4.2 .

Table 4.2 summarizes the MAPEs of estimated average density and upstream flow-rate over three time periods under different CMRs $(100 \%, 50 \%, 20 \%, 5 \%)$. The means followed by the standard deviations of the MAPEs are calculated based on the results of 100 runs each. From the results, we can see that the mean MAPE of the estimated average density is much smaller than that of the upstream flow-rate. This is due to the fact that the estimated average density only involves the estimated initial number of vehicles, which is affected by the measurement error and FIFO error. While the estimated upstream flow-rate relies on the estimated shock wave speed, which is subject to not only measurement error and FIFO error, but also modeling errors associated with Newell's simplified kinematic wave theory. As expected, the variance of MAPE is higher with smaller CMR in all cases, indicating lower precision with fewer reidentification samples. The mean MAPE of the estimated average density is higher with smaller CMR. This indicates that the bias of the average density estimation increases with lower CMR. However, the bias of the estimated upstream flow-rate is relatively constant since the mean MAPE is almost the same for all CMRs. Both estimated traffic state have smaller mean MAPEs in the second time period (8:05


Figure 4.5: Estimated upstream flow-rate v.s. observation

AM-8:20 AM). This is probably because the second dataset contains fewer errors from the sources mentioned in Section 4.2.1

Table 4.2: MAPE of estimated average density and upstream flow-rate

| Traffic State | CMR | MAPE (\%): Mean (Std dev) |  |  |
| :---: | :---: | :---: | :---: | :---: |
|  |  | $7: 50-8: 05$ | $8: 05-8: 20$ | $8: 20-8: 35$ |
| Average Density | $100 \%$ | $1.47(0.00)$ | $0.15(0.00)$ | $0.57(0.00)$ |
|  | $50 \%$ | $1.47(0.43)$ | $0.38(0.29)$ | $0.71(0.43)$ |
|  | $20 \%$ | $1.43(0.86)$ | $0.81(0.57)$ | $0.95(0.64)$ |
|  | $5 \%$ | $1.90(1.48)$ | $1.72(1.23)$ | $1.80(1.51)$ |
| Upstream Flow-rate | $100 \%$ | $13.99(0.00)$ | $10.87(0.00)$ | $14.50(0.00)$ |
|  | $50 \%$ | $14.72(0.41)$ | $12.11(0.95)$ | $14.49(0.12)$ |
|  | $20 \%$ | $14.54(0.49)$ | $11.76(0.96)$ | $14.52(0.14)$ |
|  | $5 \%$ | $14.39(0.79)$ | $11.84(1.22)$ | $14.47(0.31)$ |

### 4.4.4 A comparison with the STD method

In this subsection, we compare the proposed estimation method with the STD method in Deng et al. (2013) using the same NGSIM I-80 dataset from 5:15 PM to 5:30 PM as a benchmark. The warm-up period is set to be 5 minutes to match the setting in Deng et al. (2013). Three different CMRs.1. $1 \%, 2 \%$, and $5 \%$ were used in Deng et al. (2013), which is essentially the proportion of reidentification pairs used in the study. We estimate the average density according to (4.13) given different combinations of CMRs and flow counting intervals. Each combination is repeated 100 times, and the means and standard deviations of the MAPEs for different CMRs of reidentified vehicles and sampling intervals (the time-step size $\Delta t$ ) are shown in Figure 4.6 .

[^2]

Figure 4.6: Impact of Flow Counting Interval and CMR on Average Density Estimation

From Figure 4.6 we can see that the CMR of the reidentified vehicles has significant impacts on the estimation results: the mean and variance of the MAPE are both higher with smaller CMR. But the effect of the time counting interval is insignificant: in the case when the CMR is $1 \%$, the mean increases slightly while the variance stays almost the same with larger counting interval.

Comparing the results of Figure 4.6 with those of Deng et al. (2013) ${ }^{2}$, our method

[^3]outperforms the STD method for all scenarios. For example, when the CMR is $1 \%$, and the sampling interval is 1 min , the MAPE of the average density is about $24 \%$ by the STD method, but $8.5 \%$ by our method. Note that in Deng et al. (2013), the initial number of vehicles was assumed given and accurate, and $W$ was predefined. In contrast, we estimate $n_{0}, W$, and $K$ simultaneously. Thus, the simultaneous estimation framework presented here yields better results, yet without determining the initial number of vehicles and the shock wave speed separately.

### 4.5 Summary

This chapter proposed the LoopReid method to simultaneously estimate states and parameters with Eulerian traffic count data from loop detectors and Lagrangian vehicle trajectory data from reidentification technologies. The method was developed based on Newell's simplified kinematic wave theory, which uses cumulative flow to describe the traffic state. We first formulated an optimization problem in terms of the initial number of vehicles and model parameters for a road segment, from which other traffic state can be calculated accordingly. We showed that the optimization problem can be decoupled and can be solved either analytically or with the Gauss-Newton method. We also pointed out a potential issue under stationary traffic conditions. Finally we applied the method to the NGSIM data and demonstrated its validity based on the mean absolute percentage errors of both the average density and the upstream flow-rate. A comparison with the STD method in Deng et al. (2013) further highlighted the advantage of the simultaneous estimation framework.

## Chapter 5

## Traffic Estimation with Complete Trajectory Data

### 5.1 Introduction

A major limitation of the LoopReid method proposed in Chapter 4 is that it only works for fully congested road segments and cannot handle mixed traffic state. This is not much of a problem in urban freeways network during peak hours where the congestion pattern is relatively stable. However, this limits the application of the LoopReid model to general road network at any time. In this chapter, we propose and formulate the LoopCT estimation method, which can be viewed as an extension of the LoopReid method. The LoopCT method extends the LoopReid method to accommodate mixed traffic conditions, at the cost of using more information. While the LoopReid method requires loop detector data and vehicle reidentification data (partial vehicle trajectories), the LoopCT method required loop detector data and complete trajectory (CT) data. A detailed comparison of LoopReid method and

LoopCT method is provided later in this chapter.

The rest of the chapter is organized as follows: in Section 5.2, we derive a counterpart of the Newell's simplified kinematic wave model in the Lagrangian coordinates under Eulerian boundary conditions. Then we prove the single transition theorem in the Lagrangian coordinate, which states that vehicles moving through the road segment can encounter at most one transition from free-flow traffic condition to congested traffic condition. This splits a vehicle trajectory into two parts: the vehicle moves in free-flow (or critical) traffic before the transition and in congested (or critical) traffic after the transition. Based on this property, we formulate the LoopCT traffic estimation method as a least square estimation problem in Section 5.3. The LoopCT method is compared with the LoopReid method in details. A vehicle trajectory estimation method based on the single transition theorem is also proposed in this section. Section 5.4 discussed the solution method along with some analyses of the problem. In Section 5.5, the LoopCT method is applied to the NGSIM I80 dataset with mixed traffic state. We evaluate the performance of the estimation method by comparing the estimated average density, cumulative count contour, vehicle trajectory with the observations respectively.

### 5.2 A hybrid kinematic wave model in Lagrangian coordinates under Eulerian boundary conditions

Recall the Newell's simplified kinematic wave model reviewed in Section 2.2 .

$$
\begin{equation*}
\left.N(t, x)=\min \left\{F\left(t-\frac{x}{V}\right)+n_{0}, G\left(t-\frac{l-x}{W}\right)+K(l-x)\right)\right\}, \tag{5.1}
\end{equation*}
$$

where $N(t, x)$ is the cumulative flow at time $t$ location $x . F(t)$ and $G(t)$ are cumulative counts at upstream and downstream of the road segment respectively. The initial number of vehicles within the road segment is denoted by $n_{0} ; l$ is the length of the road segment. In the triangular fundamental diagram (Figure 2.2), $V$ is the free-flow speed, $W$ is the shock wave speed in congested traffic, and $K$ is the jam density.

We denote

$$
\begin{equation*}
N_{1}(x, t)=F\left(t-\frac{x}{V}\right)+n_{0} \tag{5.2}
\end{equation*}
$$

and

$$
\begin{equation*}
N_{2}(x, t)=G\left(t-\frac{l-x}{W}\right)+K(l-x) . \tag{5.3}
\end{equation*}
$$

Then we can restate Newell's model in (5.1) as:

$$
\begin{equation*}
N(t, x)=\min \left\{N_{1}(t, x), N_{2}(t, x)\right\} \tag{5.4}
\end{equation*}
$$

Without loss of generality, we assume both $F(t)$ and $G(t)$ are monotonically increasing in $t$. Then it is straightforward to show that both $N_{1}(t, x)$ and $N_{2}(t, x)$ are increasing functions in $t$. We can show that $N_{1}(t, x)$ and $N_{2}(t, x)$ are decreasing functions in $x$ in the following Lemma.

Lemma 5.2.1. At a time instant, both $N_{1}(x, t)$ and $N_{2}(x, t)$ are decreasing in $x$, and their partial derivative satisfies the following relation:

$$
\begin{equation*}
-K<\frac{N_{2}\left(t, x_{2}\right)-N_{2}\left(t, x_{1}\right)}{x_{2}-x_{1}} \leq-k_{c} \leq \frac{N_{1}\left(t, x_{2}^{\prime}\right)-N_{1}\left(t, x_{1}^{\prime}\right)}{x_{2}^{\prime}-x_{1}^{\prime}}<0 \tag{5.5}
\end{equation*}
$$

for any $0 \leq x_{1}<x_{2} \leq l$ and $0 \leq x_{1}^{\prime}<x_{2}^{\prime} \leq l$.

Proof. In the triangular fundamental diagram, we have

$$
\begin{equation*}
C=V k_{c}=W\left(K-k_{c}\right) . \tag{5.6}
\end{equation*}
$$

Consider the capacity constraint at upstream/downstream locations, for any $t_{1}<t_{2}$,

$$
\begin{align*}
& 0<F\left(t_{2}\right)-F\left(t_{1}\right) \leq C\left(t_{2}-t_{1}\right),  \tag{5.7a}\\
& 0<G\left(t_{2}\right)-G\left(t_{1}\right) \leq C\left(t_{2}-t_{1}\right) \tag{5.7b}
\end{align*}
$$

For any $0 \leq x_{1}<x_{2} \leq l$, we have:

$$
\begin{array}{r}
F\left(t-\frac{x_{2}}{V}\right)<F\left(t-\frac{x_{1}}{V}\right), \\
G\left(t-\frac{l-x_{1}}{W}\right)<G\left(t-\frac{l-x_{2}}{W}\right) . \tag{5.8b}
\end{array}
$$

Regarding $N_{1}(x, t)$, we have

$$
\begin{equation*}
\frac{N_{1}\left(t, x_{2}\right)-N_{1}\left(t, x_{1}\right)}{x_{2}-x_{1}}=\frac{F\left(t-\frac{x_{2}}{V}\right)-F\left(t-\frac{x_{1}}{V}\right)}{x_{2}-x_{1}}, \tag{5.9}
\end{equation*}
$$

where,

$$
\begin{equation*}
\left(F\left(t-\frac{x_{2}}{V}\right)-F\left(t-\frac{x_{1}}{V}\right)\right) \in\left[-\frac{C \cdot\left(x_{1}-x_{2}\right)}{V}, 0\right) \tag{5.10}
\end{equation*}
$$

according to (5.7a). So

$$
\begin{equation*}
\frac{N_{1}\left(t, x_{2}\right)-N_{1}\left(t, x_{1}\right)}{x_{2}-x_{1}} \in\left[-k_{c}, 0\right) \tag{5.11}
\end{equation*}
$$

Regarding $N_{2}(x, t)$, we have

$$
\begin{equation*}
\frac{N_{2}\left(t, x_{2}\right)-N_{2}\left(t, x_{1}\right)}{x_{2}-x_{1}}=\frac{G\left(t-\frac{l-x_{2}}{W}\right)-G\left(t-\frac{l-x_{1}}{W}\right)}{x_{2}-x_{1}}-K \tag{5.12}
\end{equation*}
$$

where,

$$
\begin{equation*}
\left(G\left(t-\frac{l-x_{2}}{W}\right)-G\left(t-\frac{l-x_{1}}{W}\right)\right) \in\left(0, C \cdot\left(\frac{x_{2}-x_{1}}{W}\right)\right] \tag{5.13}
\end{equation*}
$$

according to 5.7 b . So

$$
\begin{equation*}
\frac{N_{2}\left(t, x_{2}\right)-N_{2}\left(t, x_{1}\right)}{x_{2}-x_{1}} \in\left(-K,-k_{c}\right] . \tag{5.14}
\end{equation*}
$$

Based on (5.11) and (5.14), we can conclude the results stated in Lemma 5.2.1.

Lemma 5.2.1 is first proved by Rey et al. (2015) in the study of vehicle trajectory estimation. However, their proof is not rigorous since $F(\cdot)$ and $G(\cdot)$ are not differentiable everywhere. Furthermore, they derived the relation between $N_{1}(t, x)$ and $N_{2}(t, x)$ within a road segment, which can be illustrated as in Figure 5.1. Their results showed


Figure 5.1: Illustration of relation of $N_{1}(t, x)$ and $N_{2}(t, x)$
that the whole road segment can be divided into three subsegments based on the relation between $N_{1}(t, x)$ and $N_{2}(t, x)$. Upstream of the intersection (overlapping) of the two surfaces, we have $N_{1}(t, x)<N_{2}(t, x)$, the intersection part has $N_{1}(t, x)=N_{2}(t, x)$,
and downstream of intersection has $N_{1}(t, x)>N_{2}(t, x)$. Based on Newell's kinematic wave model, the cumulative count surface $N(t, x)$ is then determined by the lower envelop of the two surfaces.

In the first-in-first-out scenario, the vehicles are labeled according to their cumulative flow order. For the vehicle with cumulative flow label $i$, we denote the location of vehicle $i$ at time $t$ as $X(t, i)$. We denote $r(i)$ as the time when vehicle $i$ enters this road segment and $s(i)$ as the time when the vehicle leaves this road segment. Then we have

$$
\begin{align*}
X(r(i), i) & =0,  \tag{5.15a}\\
X(s(i), i) & =l,  \tag{5.15b}\\
N(t, X(t, i)) & =i \tag{5.15c}
\end{align*}
$$

By definition, $X(t, i)$ is the trajectory of vehicle $i$ with bounded speed:

$$
\begin{equation*}
\frac{X\left(t_{1}, i\right)-X\left(t_{2}, i\right)}{t_{1}-t_{2}} \in[0, V] \tag{5.16}
\end{equation*}
$$

for any $r(i) \leq t_{1}<t_{2} \leq s(i)$.

Define $X_{1}(t, i)$ and $X_{2}(t, i)$ as the respective inverse function of $N_{1}(t, x)$ and $N_{2}(t, x)$, i.e., $X_{1}(t, i)$ satisfies

$$
\begin{equation*}
N_{1}\left(t, X_{1}(t, i)\right)=F\left(t-\frac{X_{1}(t, i)}{V}\right)+n_{0}=i \tag{5.17}
\end{equation*}
$$

and $X_{2}(t, i)$ satisfies

$$
\begin{equation*}
N_{2}\left(t, X_{2}(t, i)\right)=G\left(t-\frac{l-X_{2}(t, i)}{W}\right)+K\left(l-X_{2}(t, i)\right)=i \tag{5.18}
\end{equation*}
$$

Since both $N_{1}(t, x)$ and $N_{2}(t, x)$ are decreasing functions in $x$ and increasing functions in $t, X_{1}(t, i)$ and $X_{2}(t, i)$ are both well-defined and monotonically increasing in $t$. Then, we have the following theorem.

Theorem 5.2.2. For given $t$, the location of the $i$-th vehicle on the road segment, $X(t, i)$, is given by

$$
\begin{equation*}
X(t, i)=\min \left\{X_{1}(t, i), X_{2}(t, i)\right\} \tag{5.19}
\end{equation*}
$$

Proof. By Newell's kinematic wave theory, we have

$$
\begin{equation*}
N(t, X(t, i))=\min \left\{N_{1}(t, X(t, i)), N_{2}(t, X(t, i))\right\} \tag{5.20}
\end{equation*}
$$

1. When $N(t, X(t, i))=N_{1}(t, X(t, i))=N_{2}(t, X(t, i))$ :

This leads to:

$$
\begin{equation*}
N_{1}(t, X(t, i))=N_{2}(t, X(t, i))=i \tag{5.21}
\end{equation*}
$$

By definition, we have

$$
\begin{equation*}
X(t, i)=X_{1}(t, i)=X_{2}(t, i) \tag{5.22}
\end{equation*}
$$

2. When $N(t, X(t, i))=N_{1}(t, X(t, i))<N_{2}(t, X(t, i))$ :

This leads to:

$$
\begin{align*}
& N_{1}(t, X(t, i))=i=N_{1}\left(t, X_{1}(t, i)\right),  \tag{5.23a}\\
& N_{2}(t, X(t, i))>i=N_{2}\left(t, X_{2}(t, i)\right) . \tag{5.23b}
\end{align*}
$$

By the monotonicity of $N_{1}(t, x)$ and $N_{2}(t, x)$ in $x$, we have

$$
\begin{equation*}
X(t, i)=X_{1}(t, i)<X_{2}(t, i) \tag{5.24}
\end{equation*}
$$

3. When $N(t, i)=N_{2}(t, X(t, i))<X_{1}(t, X(t, i))$ :

Similar to case 2, we can show that

$$
\begin{equation*}
X(t, i)=X_{2}(t, i)<N_{1}(t, i) \tag{5.25}
\end{equation*}
$$

In summary, we have

$$
\begin{equation*}
X(t, i)=\min \left\{X_{1}(t, x), X_{2}(t, i)\right\} \tag{5.26}
\end{equation*}
$$

Rey et al. (2015) also provide a proof for Theorem 5.2.2, but the proof is not rigorous since $F(\cdot)$ and $G(\cdot)$ are not differentiable everywhere. This theorem is the counterpart of Newell's model (5.4) in the Lagrangian coordinate. Now we state and prove the counter part of Lemma 5.2.1 in the Lagrangian coordinate.

Lemma 5.2.3. For given vehicle $i$, both $X_{1}(t, x)$ and $X_{2}(t, x)$ are non-decreasing, and their partial derivative satisfies the following relation:

$$
\begin{equation*}
0 \leq \frac{X_{2}\left(t_{2}, i\right)-X_{2}\left(t_{1}, i\right)}{t_{2}-t_{1}} \leq \frac{X_{1}\left(t_{2}^{\prime}, i\right)-X_{1}\left(t_{1}^{\prime}, i\right)}{t_{2}^{\prime}-t_{1}^{\prime}}=V \tag{5.27}
\end{equation*}
$$

for any $r(i) \leq t_{1}<t_{2} \leq s(i)$ and $r(i) \leq t_{1}^{\prime}<t_{2}^{\prime} \leq s(i)$.

Proof. Let $r(i) \leq t_{1}<t_{2} \leq s(i)$. For $X_{1}(t, x)$, by definition,

$$
\begin{equation*}
N_{1}\left(t_{1}, X_{1}\left(t_{1}, i\right)\right)=N_{1}\left(t_{2}, X_{1}\left(t_{2}, i\right)\right)=i \tag{5.28}
\end{equation*}
$$

This leads to

$$
\begin{equation*}
F\left(t_{1}-\frac{X_{1}\left(t_{1}, i\right)}{V}\right)=F\left(t_{2}-\frac{X_{1}\left(t_{2}, i\right)}{V}\right) \tag{5.29}
\end{equation*}
$$

Since $F(t)$ is a monotonically increasing function, we have

$$
\begin{equation*}
t_{1}-\frac{X_{1}\left(t_{1}, i\right)}{V}=t_{2}-\frac{X_{1}\left(t_{2}, i\right)}{V} \tag{5.30}
\end{equation*}
$$

which leads to

$$
\begin{equation*}
\frac{X_{1}\left(t_{2}, i\right)-X_{1}\left(t_{1}, i\right)}{t_{2}-t_{1}}=V . \tag{5.31}
\end{equation*}
$$

For $X_{2}(t, x)$, by definition,

$$
\begin{equation*}
N_{2}\left(t_{1}, X_{2}\left(t_{1}, i\right)\right)=N_{2}\left(t_{2}, X_{2}\left(t_{2}, i\right)\right)=i \tag{5.32}
\end{equation*}
$$

This leads to

$$
\begin{equation*}
\frac{X_{2}\left(t_{2}, i\right)-X_{2}\left(t_{1}, i\right)}{t_{2}-t_{1}}=\frac{G\left(t_{2}-\frac{l-X_{2}\left(t_{2}, i\right)}{W}\right)-G\left(t_{1}-\frac{l-X_{2}\left(t_{1}, i\right)}{W}\right)}{K\left(t_{2}-t_{1}\right)} \tag{5.33}
\end{equation*}
$$

By 5.7b), we have
$\left.0 \leq\left(G\left(t_{2}-\frac{l-X_{2}\left(t_{2}, i\right)}{W}\right)-G\left(t_{1}-\frac{l-X_{2}\left(t_{1}, i\right)}{W}\right)\right) \leq C\left(t_{2}-t_{1}+\frac{X_{2}\left(t_{1}, i\right)-X_{2}\left(t_{2}, i\right)}{W}\right)\right]$,
which leads to

$$
\begin{equation*}
0 \leq \frac{X_{2}\left(t_{2}, i\right)-X_{2}\left(t_{1}, i\right)}{t_{2}-t_{1}} \leq \frac{C}{K-\frac{C}{W}}=V \tag{5.34}
\end{equation*}
$$

Based on (5.31) and (5.33), we can conclude the results stated in Lemma 5.2.3.

A direct consequence of Lemma 5.2 .3 is the following single transition theorem in the Lagrangian coordinate.

Theorem 5.2.4 (Single Transition theorem). The trajectory of vehicle $i, X(t, i)$ can be divided into three sub-segments, as shown in Figure 5.2.

1. For $t \in[r(i), d(i)), X(t, i)=X_{1}(t, i)<X_{2}(t, i)$,
2. For $t \in[d(i), e(i)], X(t, i)=X_{1}(t, i)=X_{2}(t, i)$,
3. For $t \in(e(i), s(i)], X(t, i)=X_{2}(t, i)<X_{1}(t, i)$.


Figure 5.2: Representation of $X_{1}(t, i)$ and $X_{2}(t, i)$

Proof. By definition and Newell's model (5.4), $N_{1}(r(i), 0)=i \leq N_{2}(r(i), 0)$. So $X_{1}(r(i), i)=0 \leq X_{2}(r(i), i)$. Similarly, we have $X_{2}(s(i), i)=l \leq X_{1}(s(i), i)$.

Consider the fact that $X_{1}(t, i)$ and $X_{2}(t, i)$ are increasing functions in $t$ and the relationship between their subderivatives in Lemma 5.2.3, we can state that

$$
\begin{equation*}
\exists[d(i), e(i)] \subseteq[r(i), s(i)]:\left\{\forall x \in[d(i), e(i)], X_{1}(t, i)=X_{2}(t, i)\right\} . \tag{5.35}
\end{equation*}
$$

In general, we would have $d(i)=e(i)$. In the case when $d(i) \neq e(i), X_{1}(t, i)$ and
$X_{2}(t, i)$ share the same increasing rate:

$$
\begin{equation*}
\frac{X_{2}\left(t_{2}, i\right)-X_{2}\left(t_{1}, i\right)}{t_{2}-t_{1}}=\frac{X_{1}\left(t_{2}, i\right)-X_{1}\left(t_{1}, i\right)}{t_{2}-t_{1}}=V . \tag{5.36}
\end{equation*}
$$

In summary, we can separate the trajectory into three sub-segments as stated in Theorem 5.2.4.

According to the single transition theorem, the vehicle travels at free-flow speed during time $[r(i), e(i)]$ and reduces speed during time $(e(i), s(i)]$. However, it is still possible for vehicle $i$ to resume free-flow speed after the transition time $e(i)$ as illustrated in Figure 5.2 around $t=t^{\prime}$. The following corollary shows the corresponding traffic condition in the three subsegments.

Corollary 5.2.5. The trajectory of vehicle $i, X(t, i)$ can be divided into three subsegments, as shown in Figure 5.2:

1. For $t \in[r(i), d(i)), N(t, X(t, i))=N_{1}(t, X(t, i))<N_{2}(t, X(t, i))$,
2. For $t \in[d(i), e(i)], N(t, X(t, i))=N_{1}(t, X(t, i))=N_{2}(t, X(t, i))$,
3. For $t \in(e(i), r(i)], N(t, X(t, i))=N_{2}(t, X(t, i))<N_{1}(t, X(t, i))$.

Proof. Recall that $N_{1}(t, x)$ and $N_{2}(t, x)$ are both decreasing functions with respect to $x$.

1. For $t \in[r(i), d(i))$, according to Theorem 5.2.4, we have $X(t, i)=X_{1}(t, i)<$ $X_{2}(t, i)$, so

$$
\begin{aligned}
& i=N_{2}\left(t, X_{2}(t, i)\right)<N_{2}(t, X(t, i)), \\
& i=N_{1}\left(t, X_{1}(t, i)\right)=N_{1}(t, X(t, i))
\end{aligned}
$$

So we have $N(t, x)=N_{1}(t, x)<N_{2}(t, x)$.
2. For $t \in[d(i), e(i)]$, we have $X(t, i)=X_{1}(t, i)=X_{2}(t, i)$, so

$$
\begin{aligned}
& i=N_{2}\left(t, X_{2}(t, i)\right)=N_{2}(t, X(t, i)) \\
& i=N_{1}\left(t, X_{1}(t, i)\right)=N_{1}(t, X(t, i))
\end{aligned}
$$

So we have $N(t, x)=N_{1}(t, x)=N_{2}(t, x)$.
3. For $t \in(e(i), l]$, we have $X(t, i)=X_{2}(t, i)<X_{1}(t, i)$, so

$$
\begin{aligned}
& i=N_{2}\left(t, X_{2}(t, i)\right)=N_{2}(t, X(t, i)) \\
& i=N_{1}\left(t, X_{1}(t, i)\right)<N_{1}(t, X(t, i))
\end{aligned}
$$

So we have $N(t, x)=N_{2}(t, x)<N_{1}(t, x)$.

Corollary 5.2.5 states that vehicle $i$ experiences free-flow traffic condition during $[r(i), d(i))$, critical traffic condition during $[d(i), e(i)]$, and congested traffic state during ( $e(i), s(i)]$. It is important to notice that the critical traffic condition is a special case of both the free-flow and congested traffic condition, since a vehicle travels at the free-flow speed $V$ during critical traffic state, it is still possible for vehicles to resume free-flow speed in the third subsegment $t \in(e(i), s(i)]$.

### 5.3 The LoopCT estimation method

The major goal of the study is to estimate the traffic state in terms of cumulative flow within a homogeneous road segment given cumulative flow counts at both ends
and trajectory data of vehicles which traveled through the segment. Formally, the following data are given:

- cumulative flows function: $F(t)$ and $G(t)$ for $t>0$,
- and trajectory functions of $I$ vehicles: $X(t, i)$, where $i=1,2, \ldots I$.

According to Corollary 5.2.5, the trajectory of an individual vehicle can be divided into three sub-segments, with $d(i)$ and $e(i)$ being the watersheds. The second subsegment, $t \in[d(i), e(i)]$, is very special since it is in critical traffic condition, which can be described using both the free-flow and congested part of the Newell's model.

### 5.3.1 An optimization problem in initial states and parameters

Assume all vehicles follow first-in-first-out principle. For vehicle $i$, its label at the upstream should match its label at the downstream:

$$
F(r(i))+n_{0}=G(s(i))
$$

As illustrated in Figure 5.3, the vehicle's label at $t=e(i)$ can be expressed using both free-flow and congested part of the Newell's model as:

$$
\begin{aligned}
& i=F(r(i))+n_{0}=N_{1}(e(i), X(e(i), i))=F\left(e(i)-\frac{X(e(i), i)}{V}\right)+n_{0} \\
& i=G(s(i))=N_{2}(e(i), X(e(i), i))=G\left(e(i)-\frac{l-X(e(i), i)}{W}\right)+K \cdot(l-X(e(i), i))
\end{aligned}
$$

The reason we choose $t=e(i)$ instead of any other $t \in[d(i), e(i)]$ will be explained later in this section.


Figure 5.3: Illustration of the LoopCT estimation method

In summary, we have the following estimation equations for vehicle $i$ :

$$
\begin{align*}
& F(r(i))+n_{0}=G(s(i)),  \tag{5.38a}\\
& i=F(r(i))+n_{0}=N_{1}(e(i), X(e(i), i))=F\left(e(i)-\frac{X(e(i), i)}{V}\right)+n_{0},  \tag{5.38b}\\
& i=G(s(i))=N_{2}(e(i), X(e(i), i))=G\left(e(i)-\frac{l-X(e(i), i)}{W}\right)+K \cdot(l-X(e(i), i)), \tag{5.38c}
\end{align*}
$$

where 5.38 b corresponds to the free-flow and critical traffic condition during time $[r(i), e(i)], 5.38 \mathrm{c})$ corresponds to the congested traffic condition during time $(e(i), s(i)]$. Recall that the $F(t)$ and $G(t)$ are monotonically increasing, 5.38b) simplifies to:

$$
\begin{equation*}
r(i)-e(i)+\frac{X(e(i), i)}{V}=0 \tag{5.39}
\end{equation*}
$$

The equations in (5.38) are subject to different types of errors, including measurement
errors, first-in-first-out violation errors, and modeling errors. In the following, we formulate an optimization problem to minimize the sum of squared errors and fit the traffic parameters and states.

Assuming we have observed complete trajectories of totally $I$ vehicles, We can formulate an optimization problem with the following objective function to estimate $\hat{n}_{0}, \hat{V}, \hat{W}, \hat{K}, \hat{e}(1), \ldots, \hat{e}(I):$

$$
\begin{align*}
& \min Z\left(\hat{n}_{0}, \hat{V}, \hat{W}, \hat{K}, \hat{e}(1), \ldots, \hat{e}(I)\right) \\
= & \sum_{i}^{I} \xi_{i}^{2}+\sum_{i}^{I} \epsilon_{i}^{2}+\sum_{i}^{I} \sigma_{i}^{2} \\
= & {\left[F(r(i))+\hat{n}_{0}-G(s(i))\right]^{2} }  \tag{5.40}\\
& +\left[\hat{e}(i)-\frac{X(\hat{e}(i), i)}{\hat{V}}-r(i)\right]^{2} \\
& +\left[G\left(\hat{e}(i)-\frac{l-X(\hat{e}(i), i)}{\hat{W}}\right)+\hat{K}(l-X(\hat{e}(i), i))-G(s(i))\right]^{2},
\end{align*}
$$

where

$$
\begin{align*}
\xi_{i} & =F(r(i))+\hat{n}_{0}-G(s(i)),  \tag{5.41a}\\
\epsilon_{i} & =\hat{e}(i)-\frac{X(\hat{e}(i), i)}{\hat{V}}-r(i),  \tag{5.41b}\\
\sigma_{i} & =G\left(\hat{e}(i)-\frac{l-X(\hat{e}(i), i)}{\hat{W}}\right)+\hat{K}(l-X(\hat{e}(i), i))-G(s(i)) . \tag{5.41c}
\end{align*}
$$

The unknowns variables are $\left(\hat{V}, \hat{W}, \hat{K}, \hat{n}_{0}, \hat{e}(1), \ldots, \hat{e}(I)\right)$, where $(V, W, K)$ are parameters in the fundamental diagram, $\hat{n}_{0}$ is the initial condition of the road segment. Note that, according to Theorem 5.2.4, when $d(i) \neq e(i)$, any $t(i) \in[d(i), e(i)]$ satisfies (5.38). Thus, 5.38) have multiple solution of $e(i)$ in theory. However, this is a rare case in reality when traffic stays at critical states for a long period of time. Even when multiple solution exists for $e(i)$, it does not impact the estimation accuracy of
$\left(\hat{V}, \hat{W}, \hat{K}, \hat{n}_{0}\right)$, so the parameter estimation is still accurate. Another reason we choose $e(i)$ is because its physical meaning. $e(i)$ is the latest time instance that vehicle $i$ travels at the free-flow speed, thus it can be detected easily from vehicle trajectory.

### 5.3.2 Estimation of vehicle trajectories

The Newell's model still applies to this scenario, so the estimating equations in Subsection 4.2.2 can be used to derive traffic state variables. The study by Rey et al. (2015) proposed a trajectory estimation method based on Theorem 5.2.2. However, the method uses bisection method to solve the trajectory at each time step, which is computationally intensive. Here, we derive a different trajectory estimation method based on Lemma 5.2.3. The new method calculates the vehicle trajectories iteratively in a closed form at each time step, given:

- boundary cumulative flow $F(t)$ and $G(t)$,
- vehicle $i$ 's entrance and exit time $r(i)$ and $s(i)$,
- and parameters in the fundamental diagram, $V, W$ and $K$.

Theorem 5.2.2 states that the trajectory of vehicle $i, X(t, i)$ is the minimum of two virtual trajectories $X_{1}(t, i)$ and $X_{2}(t, i)$. By (5.31), the forward trajectory $X_{1}(t, i)$ is a straight line from $(r(i), 0)$ with slope $V$, so we have:

$$
\begin{equation*}
X_{1}(t, i)=r(i)+V \cdot(t-r(i)) \tag{5.42}
\end{equation*}
$$

The way to derive the backward trajectory $X_{2}(t, i)$ is illustrated in Figure 5.4:


Figure 5.4: Vehicle trajectory estimation

We first construct a grid from the downstream at each time instant with slope $-W$. Inside each grid, by 5.33 and the geometry relationship, we have:

$$
\begin{align*}
X_{2}\left(t_{2}, i\right)-X_{2}\left(t_{1}, i\right) & =\frac{G\left(\tau_{2}\right)-G\left(\tau_{1}\right)}{K},  \tag{5.43a}\\
\frac{X_{2}\left(t_{1}, i\right)-l}{t_{1}-\tau_{1}} & =-W \tag{5.43b}
\end{align*}
$$

where $\tau_{1}$ and $\tau_{2}$ are time stamps of the grids constructed from the downstream. This leads to:

$$
\begin{align*}
X_{2}\left(t_{1}, i\right) & =X_{2}\left(t_{2}, i\right)-\frac{G\left(\tau_{2}\right)-G\left(\tau_{1}\right)}{K}  \tag{5.44a}\\
t_{1} & =\frac{l-X_{2}\left(t_{1}, i\right)}{W}+\tau_{1} \tag{5.44b}
\end{align*}
$$

This result provides a way to solve the trajectory on the grid line, $\left(t_{1}, X_{2}\left(t_{1}, i\right)\right)$, given the previous trajectory point $\left(t_{2}, X_{2}\left(t_{2}, i\right)\right)$. The backward trajectory $X_{2}(t, i)$ can be constructed according to the following steps:

1. Set $\left(t_{2}, X_{2}\left(t_{2}, i\right)\right)^{(1)}=(s(i), l)$.
2. Calculate $\left(t_{1}, X_{2}\left(t_{1}, i\right)\right)^{(i)}$ based on given $\left(t_{2}, X_{2}\left(t_{2}, i\right)\right)^{(i)}$.
3. Update $\left(t_{2}, X_{2}\left(t_{2}, i\right)\right)^{(i+1)}=\left(t_{1}, X_{2}\left(t_{1}, i\right)\right)^{(i)}$. If $X_{2}\left(t_{2}, i\right)^{(i+1)} \leq 0$, stop; otherwise go back to step 2.

The backward trajectory $X_{2}(t, i)$ is then constructed as a piece-wise linear function from the points $\left(t_{2}, X_{2}\left(t_{2}, i\right)\right)^{(1)},\left(t_{2}, X_{2}\left(t_{2}, i\right)\right)^{(2)}$, and so on. The estimated trajectory of vehicle $i$ at time $t$ is thus $X(t, i)=\min \left\{X_{1}(t, i), X_{2}(t, i)\right\}$ as shown in Figure5.4.

### 5.3.3 A comparison between LoopReid method and LoopCT method

The LoopCT method proposed in this chapter can be viewed as an extension of the LoopReid method proposed in Chapter 4. Recall the estimation equations in LoopReid method,

$$
\begin{align*}
F(r(i))+n_{0} & =G(s(i))  \tag{5.45a}\\
G(s(i)) & =G\left(r(i)-\frac{l}{W}\right)+K l \tag{5.45b}
\end{align*}
$$

and in LoopCT method,

$$
\begin{align*}
F(r(i))+n_{0} & =G(s(i)),  \tag{5.46a}\\
F(r(i))+n_{0} & =F\left(e(i)-\frac{X(e(i), i)}{V}\right)+n_{0},  \tag{5.46b}\\
G(s(i)) & =G\left(e(i)-\frac{l-X(e(i), i)}{W}\right)+K \cdot(l-X(e(i), i)) . \tag{5.46c}
\end{align*}
$$

(5.45a) is identical to 5.46a. When the road segment is fully congested, we have $e(i)=r(i)$, where the transition time aligns with the entrance time for all vehicle. (5.46c) degenerates to (5.45b), and 5.46b) degenerates to $0=0$. In this case, the two methods have the same set of estimation equations. As a side note, this phenomenon is also verifies using the NGSIM dataset. The LoopCT method is implemented on I101 dataset used in Section 4.4. We found $e(i)$ equals $r(i)$ for all vehicles when the free-flow speed is predetermine to be 60 mph . Thus we get identical result as the LoopReid method for the same dataset.

In addition, the degenerated formulations are free from the term, $X(e(i), i)$, so only the vehicle entrance and exit time are relevant. In this sense, the LoopCT method extends the LoopReid method to mixed traffic state, at the cost of using more complete vehicle trajectory data (reidentification data v.s. complete trajectory).

### 5.4 Solution and Optimization Method

### 5.4.1 A decoupling method with predetermined free-flow speed

Following a similar fashion of Theorem 4.3.1, the objective function 5.40 can also be decoupled, where we minimize $\sum_{i}^{I} \xi_{i}^{2}$ and $\sum_{i}^{I} \epsilon_{i}+\sum_{i}^{I} \sigma_{i}^{2}$ in parallel.

[^4]$\sum_{i}^{I} \xi_{i}^{2}$ only involves $\hat{n}_{0}$ and the optimal can be found by taking the average according Corollary 4.3.2. Theoretically, $\hat{n}_{0}$ can be solved using only one trajectory. The minimization of $\sum_{i}^{I} \epsilon_{i}+\sum_{i}^{I} \sigma_{i}^{2}$ involves $(\hat{V}, \hat{W}, \hat{K}, \hat{e}(1), \ldots, \hat{e}(I)$. With $I$ trajectories, the number of unknown is $I+3$, the number of equations is $2 I$. So it is solvable with more than three vehicle trajectories.

A challenge with minimizing of $\sum_{i}^{I} \epsilon_{i}+\sum_{i}^{I} \sigma_{i}^{2}$ is that the number of variables increase linearly with the number of vehicle trajectories in consideration, which can be cumbersome to solve with large data size. According to the physical insights revealed in Theorem 5.2.4 $e(i)$ is the last time instant when vehicle $i$ moves with free-flow speed. By examining typical vehicle trajectory data in the NGSIM dataset, we found three types of trajectories as illustrated in Figure 5.5, where the dash line has the slope of free-flow speed.

In Figure 5.5, trajectory 1 represents vehicles which entered the segment at $r(i)$ and left at $s_{1}(i)$ with free-flow speed throughout the whole segment. Trajectory 2 represents a vehicle entered the segment with free-flow speed at $r(i)$ and then left with reduced speed at $s_{2}(i)$ due to congestion. Trajectory 3 stands for a vehicle traveled in congestion throughout the whole segment from $r(i)$ to $s_{3}(i)$. We assume $V$ is predetermined, which should not be a problem in practice as $V$ is usually the speed limit at the road segment. To find $e(i)$ for vehicle $i$ given $V$, we first find

$$
\begin{equation*}
e^{\prime}(i)=\{t: X(t, i)=V(t-r(i)), t \in(r(i), s(i)]\} \tag{5.47}
\end{equation*}
$$

using the bisection method. Then, $e(i)$ can be determined according to the following


Typical Vehicle Trajectory $\qquad$
Virtual Trajectory with Free-flow Speed -------

Figure 5.5: Typical vehicle trajectories
conditions, which corresponds to the three types of trajectories in Figure 5.5.

$$
e(i)=\left\{\begin{array}{ll}
s(i), & e^{\prime}(i) \text { does not exist, } s(i)-r(i) \leq \frac{l}{V} ;(\text { Type 1) }  \tag{5.48}\\
e^{\prime}(i), & e^{\prime}(i) \text { exists; (Type 2) } \\
r(i), & e^{\prime}(i) \text { does not exist, } s(i)-r(i)>\frac{l}{V} ;(\text { Type } 3)
\end{array} .\right.
$$

By predetermining $e(i)$ given $V$, the problem is greatly simplified because the number of variables becomes irrelevant to the data size. The simplified optimization problem
is:

$$
\begin{align*}
& \min Z\left(\hat{n}_{0}, \hat{W}, \hat{K},\right) \\
= & \sum_{i}^{I} \xi_{i}^{2}+\sum_{i}^{I} \sigma_{i}^{2}  \tag{5.49}\\
= & {\left[F(r(i))+\hat{n}_{0}-G(s(i))\right]^{2}+} \\
& {\left[G\left(e(i)-\frac{l-X(e(i), i)}{\hat{W}}\right)+\hat{K}(l-X(e(i), i))-G(s(i))\right]^{2} . }
\end{align*}
$$

The decoupling method can be used to minimize $\xi_{i}^{2}$ and $\sigma_{i}^{2}$ respectively.
To solve $\min \sum_{i}^{I} \xi_{i}^{2}$, the optimal is shown in Corollary 4.3.2.

$$
\begin{equation*}
\hat{n}_{0}=\sum_{i}^{I} \frac{[F(r(i))-G(s(i))]}{I} . \tag{5.50}
\end{equation*}
$$

To solve $\min \sum_{i}^{I} \xi_{i}^{2}$, we can use a similar Gaussian Newton method as described in Section 4.3. We denote $\boldsymbol{\theta}$ as the column vector of $(\hat{W}, \hat{K})^{T}$. Starting from an initial guess $\boldsymbol{\theta}^{(0)}=\left(\hat{W}^{(0)}, \hat{K}^{(0)}\right)^{T}$, the Gauss-Newton method updates the results by iterating

$$
\boldsymbol{\theta}^{(j+1)}=\boldsymbol{\theta}^{(j)}-\left[\boldsymbol{J}\left(\boldsymbol{\theta}^{(j)}\right)^{T} \boldsymbol{J}\left(\boldsymbol{\theta}^{(j)}\right)\right]^{-1} \boldsymbol{J}\left(\boldsymbol{\theta}^{(j)}\right)^{T} \boldsymbol{\xi}\left(\boldsymbol{\theta}^{(j)}\right)
$$

where $\boldsymbol{J}(\boldsymbol{\theta})$ is the Jacobian matrix of $\boldsymbol{\xi}(\boldsymbol{\theta})=\left(\xi_{1}(\boldsymbol{\theta}), \xi_{2}(\boldsymbol{\theta}), \ldots, \xi_{I}(\boldsymbol{\theta})\right)^{T}$ and is defined as

$$
\boldsymbol{J}(\boldsymbol{\theta})=\left[\begin{array}{cc}
\frac{\partial \xi_{1}(\boldsymbol{\sigma})}{\partial \hat{W}} & \frac{\partial \xi_{1}(\boldsymbol{\sigma})}{\partial \tilde{K}}  \tag{5.51}\\
\vdots & \vdots \\
\frac{\partial \xi_{I}(\boldsymbol{\sigma})}{\partial \hat{W}} & \frac{\partial \xi_{I}(\boldsymbol{\sigma})}{\partial \hat{K}}
\end{array}\right]
$$

where,

$$
\begin{align*}
\frac{\partial \sigma}{\partial \hat{W}} & =G^{\prime}\left(e(i)-\frac{l-X(e(i), i)}{\hat{W}}\right)\left[\frac{l-X(e(i), i)}{\hat{W}^{2}}\right]  \tag{5.52}\\
\frac{\partial \sigma}{\partial \hat{K}} & =l-X(e(i), i)-G(s(i)) . \tag{5.53}
\end{align*}
$$

The downstream flow-rate in 5.52, $G^{\prime}\left(e(i)-\frac{l-X(e(i), i)}{W}\right)$, is approximated by the the forward finite difference,

$$
\begin{equation*}
G^{\prime}\left(e(i)-\frac{l-X(e(i), i)}{\hat{W}}\right) \approx \frac{G\left(e(i)-\frac{l-X(e(i), i)}{\hat{W}}+\Delta t\right)-G\left(e(i)-\frac{l-X(e(i), i)}{\hat{W}}\right)}{\Delta t}, \tag{5.54}
\end{equation*}
$$

where $\Delta t$ is the time-step size. For simplicity, we fix the time-step size to be 30 seconds.

### 5.4.2 Potential issue with near-stationary state traffic

In the case of stationary state traffic when the congestion queue tail has zero speed, all vehicles have similar trajectories as shown in Figure 5.6. By (5.46c), we have

$$
\begin{equation*}
K=\frac{G\left(e(i)-\frac{l-X(e(i), i)}{W}\right)-G(s(i))}{l-X(e(i), i)} \tag{5.55}
\end{equation*}
$$

where $l-X(e(i), i)$ and $s(i)-e(i)$ are constant for any vehicle $i$. If we define $g$ as the average downstream flow rate during time $\left[e(i)-\frac{l-X(e(i), i)}{W}, s(i)\right]$ with is a constant, 5.55) can be simplified as

$$
\begin{equation*}
K=g \cdot\left(\frac{s(i)-e(i)}{l-X(e(i), i)}+\frac{1}{W}\right) . \tag{5.56}
\end{equation*}
$$



Typical Vehicle Trajectory

Figure 5.6: Stationary state with constant queue

This result implies that the parameter $W$ and $K$ follows a hyperbolic relationship in the stationary state traffic. The estimation problem is ill-posed in this case as $K$ and $W$ cannot be uniquely determined. This is similar to the case discussed in Subsection 4.3.2.

### 5.5 Application to NGSIM I80 dataset

The I-80 NGSIM dataset is used to test the proposed estimation method as this dataset starts with free-flow traffic and gets into congested traffic in the last 10 minutes. The dataset is collected on Freeway I-80 between 2:35 pm to 3:05 pm. A homogeneous road segment with five lanes illustrated in is used in this study. The rightmost lane is excluded in the study due to excessive lane changing and queuing caused by on/off ramps. The study site is illustrated in Figure 5.7

The data collection frequency is 15 HZ . There are 4733 trajectories used in this data


Figure 5.7: Illustration of I-80 study site
set. We used 2-min warm-up time at the beginning of time period to ensure all vehicle trajectories are fully captured.

### 5.5.1 Data Preparation

We deploy virtual detectors at both ends of the road segments and generate volume counts and vehicle entry/exit times based on vehicles' longitudinal coordinates ("LocalY" field in the NGSIM trajectory dataset). We use the same method as discussed in Subsection 4.4.1 to derive cumulative flow count data at upstream/downstream locations.

The NGSIM vehicle trajectory data are collected in the form of vehicle location at
each sampling time. Due to the measurement errors, the observed vehicle location scatters around the actual unknown vehicle path. To make the errors unbiased, it is possible to project the points over the lane alignment. This is the reason we use the value "LocalY" in the NGSIM dataset to measure the distance traveled, which are the longitudinal coordinate of the front center of the vehicles. For simplicity, we use linear interpolation to reconstruct the longitudinal distance traveled by each vehicle based on "LocalY". Formally, we have

$$
\begin{equation*}
X(t, i)=X(j \Delta t, i)+\frac{X((j+1) \Delta t, i)-X(j \Delta t, i)}{\Delta t} \cdot(t-j \Delta t), \tag{5.57}
\end{equation*}
$$

where $j$ satisfies $j \Delta t \leq t \leq(j+1) \Delta t$.

### 5.5.2 Estimation of initial states and model parameters

We estimate the initial states and model parameters under the four different market penetration rates (MPR) $(100 \%, 50 \%, 20 \%$ and $5 \%)$ which is also used when validating LoopReid method in Section 4.4. In other words, we assume only the corresponding percentage of vehicle trajectories are observed. The same 2-minutes warm-up period is used to ensure all vehicle trajectories are fully tracked in the study domain. We implemented the solution method as in 5.4. The solution method requires the free-flow speed to be given in advance. To test the sensitivity of the free-flow speed, we used four different values of free-flow speeds ( $55 \mathrm{mph}, 60 \mathrm{mph}, 65 \mathrm{mph}$, and 70 mph ). The initial guess of $\hat{W}$ is 20 mph and that of $\hat{K}$ is 200 vehicles per mile per lane for all cases.

The results in Table 5.1 illustrates the mean value of $\hat{n}_{0}, \hat{W}$, and $\hat{K}$ followed by their standard deviation for 100 runs each. The true value of the initial number of vehicles, $n_{0}$, is shown in the parentheses in the header line.

The initial states is estimated with high accuracy. Even in the case with $5 \%$ MRP, the error is around $1 \%$ with subtle standard deviation. The mean values of the parameters are relatively constant across different MPRs given the same free-flow speed. The standard deviation of the parameter estimation is much larger for the case with lower MPR. The estimated shock wave speed is larger for cases with larger free-flow speed, while the estimated jam density is smaller.

Table 5.1: Parameter estimation

| Case No. | V (mph) | MPR (\%) | $\hat{n}_{0}$ (True: 38) | $\hat{W}(m p h)$ | K (vpmpl) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | 50.00 | 100.00 | 38.00 (0.00) | 15.24 (0.00) | 205.57 (0.00) |
| 2 |  | 50.00 | 38.03 (0.18) | 15.30 (0.46) | 205.30 (4.13) |
| 3 |  | 20.00 | 38.04 (0.38) | 15.74 (0.84) | 203.11 (8.82) |
| 4 |  | 5.00 | 37.83 (0.42) | 15.23 (2.02) | 206.14 (12.66) |
| 5 | 55.00 | 100.00 | 38.00 (0.00) | 16.25 (0.00) | 192.43 (0.00) |
| 6 |  | 50.00 | 37.98 (0.10) | 15.76 (0.90) | 196.57 (6.92) |
| 7 |  | 20.00 | 37.98 (0.18) | 16.03 (1.51) | 194.90 (12.00) |
| 8 |  | 5.00 | 37.96 (0.37) | 16.33 (4.71) | 196.01 (18.00) |
| 9 | 60.00 | 100.00 | 38.00 (0.00) | 21.63 (0.00) | 166.73 (0.00) |
| 10 |  | 50.00 | 37.99 (0.10) | 19.15 (5.22) | 173.88 (23.84) |
| 11 |  | 20.00 | 37.99 (0.16) | 18.38 (3.41) | 175.49 (18.28) |
| 12 |  | 5.00 | 38.05 (0.45) | 20.37 (9.04) | 173.47 (31.77) |
| 13 | 65.00 | 100.00 | 38.00 (0.00) | 34.31 (0.00) | 130.64 (0.00) |
| 14 |  | 50.00 | 38.00 (0.09) | 33.35 (10.66) | 123.51 (27.20) |
| 15 |  | 20.00 | 38.01 (0.17) | 31.04 (9.87) | 128.26 (26.03) |
| 16 |  | 5.00 | 38.17 (0.30) | 37.20 (78.62) | 125.55 (39.00) |

The objective function corresponding to case 9 is shown in Figure 5.8 as a heat map. We pick 900 points from the parameter space of $\hat{W} \in[10,40]$ and $\hat{K} \in[50,300]$ with 30 grids along each axis. The trace of the iterations is also shown in the plot. Started from the initial guess of $\left(\hat{W}^{0}, \hat{K}^{0}\right)=(20,200)$, the algorithm found the optimal in 8 iterations. The contour lines roughly follow hyperbolic relationships, which implies the existence of near-stationary traffic state in reality as derived in 5.56. Comparing the heatmap in Figure 5.8 with that in Figure 4.4 , the objective function in Figure 5.8 seems steeper. This is probably because the I80 data contains both free-flow and congested traffic, thus has more fluctuations than the I101 dataset.


Figure 5.8: Objective function plot with trace

### 5.5.3 Validation using average density and cumulative count contour

In this subsection, we compared the estimated average density, cumulative flow contour with their corresponding observations.

The average density within the segment is estimated using the (4.13) for every second. The mean absolute percentage error (MAPE) defined in (4.27) is used to evaluate the performance of the average density estimation. The cumulative count contour is calculated based on (4.11) at the resolution of 20 ft by 10 second. Since the absolute value of cumulative count depends on the choice of reference vehicle, the use of MAPE as a metric for cumulative count estimation is not meaningful. Thus, we use the mean absolute error (MAE) instead. The MAE of the cumulative count contour is defined as:

$$
\begin{equation*}
M A E_{N}=\frac{\sum_{j=1}^{J} \sum_{i=1}^{I}|\hat{N}(i \Delta t, j \Delta x)-N(i \Delta t, j \Delta x)|}{I J} \tag{5.58}
\end{equation*}
$$

where $\Delta t=10 \mathrm{~s}$ and $\Delta x=20 \mathrm{ft}$ as mentioned before. Both average density and cumulative contour are calculated for 100 runs for each combination of free-flow speed ( $50 \mathrm{mph}, 55 \mathrm{mph}, 60 \mathrm{mph}$, and 65 mph ) and MPR ( $100 \%, 50 \%, 20 \%, 5 \%$ ). The average of the MAPE for average density estimate and MAE for cumulative contour estimation are shown in Table 5.2 with the corresponding standard deviated listed in the parentheses.

Table 5.2: MAPE/MAE of estimated average density and cumulative contour

| Traffic State | MPR | $\mathrm{V}=50 \mathrm{mph}$ | $\mathrm{V}=55 \mathrm{mph}$ | $\mathrm{V}=60 \mathrm{mph}$ | $\mathrm{V}=65 \mathrm{mph}$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Average Density <br> (MAPE\%) | $100 \%$ | $0.00(0.00)$ | $0.00(0.00)$ | $0.00(0.00)$ | $0.00(0.00)$ |
|  | $50 \%$ | $0.24(0.18)$ | $0.29(0.22)$ | $0.23(0.24)$ | $0.26(0.16)$ |
|  | $20 \%$ | $0.54(0.35)$ | $0.54(0.36)$ | $0.45(0.34)$ | $0.48(0.37)$ |
|  | $5 \%$ | $1.34(0.89)$ | $1.17(0.61)$ | $1.20(0.93)$ | $0.99(0.68)$ |
| Cumulative Counts <br> (MAE) | $100 \%$ | $2.35(0.00)$ | $2.28(0.00)$ | $2.27(0.00)$ | $3.16(0.00)$ |
|  | $50 \%$ | $2.33(0.02)$ | $2.28(0.02)$ | $2.28(0.03)$ | $2.90(0.02)$ |
|  | $20 \%$ | $2.34(0.06)$ | $2.29(0.05)$ | $2.20(0.07)$ | $2.94(0.14)$ |
|  | $5 \%$ | $2.35(0.05)$ | $2.28(0.09)$ | $2.27(0.09)$ | $3.06(0.26)$ |



Figure 5.9: Absolute difference between the estimated and observed cumulative count

The performance of the average density estimation seems to be highly related with the MPR. However, even under low MPR (5\%), the estimation is still reliable with
mean MAPE around $1 \%$ and standard deviation lower than $1 \%$. The value of the free-flow speed has limited impact on the MAPE, but we can still tell $V=60$ seems to be a good choice of free-flow speed for this dataset.

The MPR has little impact on the cumulative count estimation. The mean MAE seems to be constant for different MPR with the same free-flow speed. The standard deviation of MAE increases with lower MPR. However, the effect is not very significant. The cases with free-flow speed of 60 mph has the lowest average MAE and standard deviation, which implies that the optimal free-flow speed should be around 60 mph .


Figure 5.10: Observed density contour plot

The absolute difference between estimated cumulative contour and the observation for different predetermined free-flow speeds are shown in Figure 5.9. By referring to the observed density contour plot in Figure 5.10, we found that the cumulative count estimation is more accurate when the density is low and most of the errors appears in the last 10 minutes when traffic is fully congested.

### 5.5.4 Vehicle trajectory estimation

The trajectory estimation method proposed in Subsection 5.3.2 is implemented here. The estimation error is calculated based on the difference between the observed and estimated cumulative trajectory. Formally, we define the estimation error of vehicle $i$ as:

$$
\begin{equation*}
\operatorname{Error}(i)=\frac{A_{1}(i)}{A_{2}(i)}=\frac{\int_{r(i)}^{s(i)}|\hat{X}(t, i)-X(t, i)| d t}{\int_{r(i)}^{s(i)}|X(t, i)| d t} \approx \frac{\sum_{j=\frac{r(i)}{\Delta t}}^{\frac{s(i)}{\Delta t}}|\hat{X}(j \Delta t, i)-X(j \Delta t, i)|}{\sum_{j=\frac{r(i)}{\Delta t}}^{\frac{s(i)}{\Delta t}}|X(j \Delta t, i)|} \tag{5.59}
\end{equation*}
$$

where $A_{1}(i)$ is the absolute difference between the observed and estimated cumulative trajectory, $A_{2}(i)$ is the observed cumulative trajectory. $\Delta t$ is the time step, which is fixed to be 10 seconds in this case.

Figure 5.11 provides the trajectory estimation of vehicle $\# 3164$ compared with observation. The calculated areas $A_{1}$ and $A_{2}$ are also shown in the figure.


Figure 5.11: Estimation v.s. observation, Vehicle \#3164

We tested four different sets of parameters with free-flow speed of $50 \mathrm{mph}, 55 \mathrm{mph}, 60$ mph, and 65 mph under $100 \%$ MPR. The errors are calculated for each case with 100 runs and the mean and standard deviation of the errors are summarized in Table 5.3,

Table 5.3: Trajectory estimation error

| Case No. | V (mph) | $\mathrm{W}(\mathrm{mph})$ | $\mathrm{K}(\mathrm{vpmpl})$ | mean (\%) | sd (\%) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | 50 | 15.24 | 205.57 | 4.85 | 5.60 |
| 2 | 55 | 16.25 | 192.43 | 5.06 | 5.85 |
| 3 | 60 | 21.63 | 166.73 | 5.76 | 6.35 |
| 4 | 65 | 34.31 | 130.64 | 6.89 | 6.24 |

The cases with free-flow speed 50 mph have the lowest mean and standard deviation of errors. Recall that the parameters estimated with free-flow speed 60 mph yields the best cumulative contour and average density estimation. It is most likely that there is not a uniformly best set of parameters that works for all traffic state estimation.

### 5.6 Summary

This chapter proposed another traffic state estimation method, the LoopCT method. As an extension of the LoopReid method proposed in Chapter 4, the LoopCT method can handle both congested and free-flow traffic. This chapter proved several lemmas and theorems, which provided important insights about the traffic dynamic within a homogeneous link. In particular, the single transition theorem leads to a new method for trajectory estimation based on boundary flux and vehicle entrance/exit time. We also showed that the LoopCT method degenerates to the LoopReid method when the road segment is fully congested using both theoretical analysis and data analysis. By predetermining the free-flow speed, the original estimation problem is greatly simplified and the gradient-based Gauss-Newton method can be used to find the optimal set. In the end, we implement our method using the NGSIM I80 dataset which contains
both free-flow and congested traffic. We achieved considerable accuracy for average density estimation, cumulative contour estimation, and vehicle trajectory estimation.

## Chapter 6

## Conclusion

### 6.1 Summary

A complete traffic estimation framework consists of traffic data, traffic flow model, and the traffic estimation method. This dissertation presents a new traffic estimation framework to reconstruct the traffic states and parameters within a homogeneous road segment with Lagrangian and Eulerian sensor data.

The dissertation started with formulating the network sensor health problem (NSHP), from which we calculate health indices for all sensors. Based on the network flow conservation principle, flows on non-base links can be derived from those on the base links. However, in reality, the network flow conservation principle can be violated due to the existence of unhealthy sensors. Thus we propose to identify an optimal set of sensors by solving an optimization problem, in which we minimize the inconsistency between derived and observed link flows. We then defined the health index of a sensor as its frequency that it appears in the optimal sensor sets. Other than finding all the optimal sets using brute-force enumeration, we present a greedy search algorithm
to find a subset of optimal sets effectively. The proposed method is applied to a road network with 30 links, among which 18 links are monitored with loop detectors. Using traffic count data from the Caltrans Performance Measurement System (PeMS) database, we show that the generated health index matches the PeMS health status very well. Although flow conservation is just a necessary (and not sufficient) condition for sensors to be healthy, this method is still quite effective and powerful. Compared with a statistic-based method, to use network flow conservation does not require predetermining a "threshold". The same method can be implemented in almost all traffic networks without calibration efforts.

The other important contributions of this dissertation are the simultaneous traffic parameter and state estimation methods for a homogeneous freeway segment proposed in Chapter 4 and Chapter 5.

Both methods share a similar optimization formulation, from which one can analytically solve the initial number of vehicles and numerically calculate model parameters. The LoopReid method incorporated loop detector data and vehicle reidentification data, but only applies to congested traffic. Via theoretically analysis, we find a potential observability problem when the shock wave speed and the jam density can not be uniquely determined under absolute stationary traffic. This issue was also mentioned in existing car-following model studies. The LoopReid method is tested using the I101 data where traffic is fully congested in the road segment under different vehicle correct matching rates. By comparing the estimated density and upstream flow rate with the observation, we show that the LoopReid method is capable to yield accurate traffic parameters and states and outperforms the benchmark stochastic three detector method in all scenarios.

Although the LoopCT method resembles a similar simultaneous parameter and state estimation framework as the LoopReid method, it adapts to both congested and
free-flow traffic. However, the LoopCT method require complete vehicle trajectory information instead of vehicle reidentification data compared with the LoopReid method. While developing the underlying traffic flow theory of the LoopCT method, we proved the single transition theorem which states that a vehicle going through a homogenous road segment can experience at most one transition from free-flow to congestion. This finding leads to a new method for vehicle trajectory estimation given upstream/downstream cumulative flow and the vehicle entrance/exit time. By predetermining free-flow speed, the optimization problem is greatly simplified and a gradient search method can be implemented to find the optimal traffic parameters. We find a similar observability issue with the traffic parameters when traffic is absolutely stationary. We used the NGSIM I80 dataset to test the performance of the method by comparing the estimated average density, cumulative flow contour, and the vehicle trajectory with the observations respectively. The LoopCT method is capable to estimate traffic state with considerable level of accuracy in most cases.

### 6.2 Future research directions

A limitation of the network sensor health problem is that it only works when at least one base link set is observable. It would be preferred if the current method can be extended to partial observable networks, which releases the constraint that the best sensor set has to achieve full link observability. Furthermore, the NSHP for large networks can be partitioned into sub-problems both in space (sub-network) and time (certain time period, e.g. peak hours). It would be interesting to explore the possibility to synthesize the results from each sub-problem and provide more insights.

The performance of the current traffic estimation methods is affected by various error sources. It would be interesting to develop a metric to quantify the amount
of errors caused by difference sources. The first-in-first-out (FIFO) is an important assumption while deriving both estimation methods. However, it is almost surely violated in reality for multi-lane traffic. We should be able to achieve higher accuracy if the FIFO violation can be incorporated into the current methods. In addition, the current methods can be extended to consider more general traffic sensor data, for example, incomplete vehicle trajectory and measurements from mid-block traffic sensor. Further research can also extend the proposed methods to network level by incorporating junction models.

## Bibliography

Aoude, G. S., Desaraju, V. R., Stephens, L. H., How, J. P., 2011. Behavior classification algorithms at intersections and validation using naturalistic data. In: Intelligent Vehicles Symposium (IV), 2011 IEEE. IEEE, pp. 601--606.

Batchelor, G. K., 2000. An introduction to fluid dynamics. Cambridge university press.

Cassidy, M. J., 1998. Bivariate relations in nearly stationary highway traffic. Transportation Research Part B 32, 49--59.

Castillo, E., Gallego, I., Sánchez-Cambronero, S., Rivas, A., 2010. Matrix tools for general observability analysis in traffic networks. Intelligent Transportation Systems, IEEE Transactions on 11 (4), 799--813.

Cetin, M., Comert, G., 2006. Short-term traffic flow prediction with regime switching models. Transportation Research Record: Journal of the Transportation Research Board (1965), 23--31.

Charnes, A., Frome, E., Yu, P.-L., 1976. The equivalence of generalized least squares and maximum likelihood estimates in the exponential family. Journal of the American Statistical Association 71, 169--171.

Chen, C., Kwon, J., Rice, J., Skabardonis, A., Varaiya, P., 2003. Detecting errors and imputing missing data for single-loop surveillance systems. Transportation Research Record: Journal of the Transportation Research Board 1855, 160--167.

Chen, C., Petty, K., 2001. Freeway performance measurement system: mining loop detector data. ... Record: Journal of the ... (01), 96--102.
URL http://trb.metapress.com/index/4Q21170168673231.pdf
Choi, J., 2014. Estimating Emissions by Modeling Freeway Vehicle Speed Profiles Using Point Detector Data. University of California, Irvine.

Claudel, C. G., Bayen, A. M., 2010a. Lax--hopf based incorporation of internal boundary conditions into hamilton--jacobi equation. part i: Theory. Automatic Control, IEEE Transactions on 55, 1142--1157.

Claudel, C. G., Bayen, A. M., 2010b. Lax--hopf based incorporation of internal boundary conditions into hamilton-jacobi equation. part ii: Computational methods. Automatic Control, IEEE Transactions on 55, 1158--1174.

Coifman, B., 2006. Vehicle level evaluation of loop detectors and the remote traffic microwave sensor. Journal of Transportation Engineering 132 (3), 213--226.

Coifman, B., Cassidy, M., 2002. Vehicle reidentification and travel time measurement on congested freeways. Transportation Research Part A 36, 899--917.

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 (4), 269--287.

Daganzo, C. F., 1997. Fundamentals of transportation and traffic operations. pp. 26--27.

Daganzo, C. F., 2005a. A variational formulation of kinematic waves: basic theory and complex boundary conditions. Transportation Research Part B 39, 187--196.

Daganzo, C. F., 2005b. A variational formulation of kinematic waves: Solution methods. Transportation Research Part B: Methodological 39 (10), 934--950.

Deng, W., Lei, H., Zhou, X., 2013. Traffic state estimation and uncertainty quantification based on heterogeneous data sources: A three detector approach. Transportation Research Part B 57, 132--157.

Dougherty, M. S., Cobbett, M. R., 1997. Short-term inter-urban traffic forecasts using neural networks. International journal of forecasting 13 (1), 21--31.

Evans, L., 2010. Partial Differential Equations. American Mathematical Society.
Gan, L., Yang, H., Wong, S. C., 2005. Traffic counting location and error bound in origin-destination matrix estimation problems. Journal of transportation engineering 131 (7), 524--534.

Gentili, M., Mirchandani, P. B., 2011. Survey of models to locate sensors to estimate traffic flows. Transportation Research Record: Journal of the Transportation Research Board 2243, 108--116.

Gohberg, I., Kailath, T., Olshevsky, V., 1995. Fast gaussian elimination with partial pivoting for matrices with displacement structure. Mathematics of Computation 64 (212), 1557--1576.

Greenshields, B., Bibbins, J., 1935. A study of traffic capacity. Highway Research. URL http://trid.trb.org/view. aspx?id=120649

Hall, F. L., Allen, B. L., Gunter, M. A., 1986. Empirical analysis of freeway flow-density relationships. Transportation Research Part A: General 20 (3), 197--210.

Herrera, J. C., Bayen, A. M., 2008. Traffic flow reconstruction using mobile sensors and loop detector data. In: Transportation Research Board 87th Annual Meeting.

Hoogendoorn, S., Van Zuylen, H., Schreuder, M., Gorte, B., Vosselman, G., 2003. Microscopic traffic data collection by remote sensing. Transportation Research Record: Journal of the Transportation Research Board (1855), 121--128.

Hu, S., Peeta, S., Chu, C., 2009. Identification of vehicle sensor locations for link-based network traffic applications. Transportation Research Part B 43 (8), 873--894.

Jeng, S.-T., 2007. Real-time vehicle reidentification system for freeway performance measurements.

Jin, W., Li, L., 2007. First-in-first-out is violated in real traffic. In: Proceedings of Transportation Research Board Annual Meeting.

Jin, W.-L., 2013. A multi-commodity lighthill-whitham-richards model of lanechanging traffic flow. Transportation Research Part B 57, 361--377.

Jin, W.-L., 2015. Continuous formulations and analytical properties of the link transmission model. Transportation Research Part B: Methodological 74, 88--103.

Jun, M., Ying, M., 2008. Research of traffic flow forecasting based on neural network. In: Intelligent Information Technology Application, 2008. IITA'08. Second International Symposium on. Vol. 2. IEEE, pp. 104--108.

Kalman, R. E., 1960. A new approach to linear filtering and prediction problems. Journal of basic Engineering 82 (1), 35--45.

Lay, D. C., 1997. Linear Algebra and Its Applications.
Lee, H., Coifman, B., 2011. Identifying and correcting pulse-breakup errors from freeway loop detectors. Transportation Research Record: Journal of the Transportation Research Board 2256, 68--78.

Lighthill, M. J., Whitham, G. B., 1955. On kinematic waves. ii. a theory of traffic flow on long crowded roads. In: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. Vol. 229. The Royal Society, pp. 317--345.

Milliken, G. A., 1990. Nonlinear regression analysis and its applications. Technometrics 32, 41--43.

Moorthy, C., Ratcliffe, B., 1988. Short term traffic forecasting using time series methods. Transportation planning and technology 12 (1), 45--56.

Muñoz, L., Sun, X., Horowitz, R., Alvarez, L., 2003. Traffic density estimation with the cell transmission model. In: American Control Conference, 2003. Proceedings of the 2003. Vol. 5. IEEE, pp. 3750--3755.

Nanthawichit, C., Nakatsuji, T., Suzuki, H., 2003. Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway. Transportation Research Record 1855 (1), 49--59.

Newell, G. F., 1993. A simplified theory of kinematic waves in highway traffic, part i: General theory. Transportation Research Part B 27, 281--287.

Ng, M., 2012. Synergistic sensor location for link flow inference without path enumeration: A node-based approach. Transportation Research Part B 46 (6), 781--788.

Oh, C., Ritchie, S. G., Oh, J.-S., Jayakrishnan, R., 2002. Real-time origin-destination (od) estimation via anonymous vehicle tracking. In: Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference on. IEEE, pp. 582--586.

Oh, C., Tok, A., Ritchie, S. G., 2005. Real-time freeway level of service using inductive-signature-based vehicle reidentification system. Intelligent Transportation Systems, IEEE Transactions on 6, 138--146.

Papageorgiou, M., Blosseville, J.-M., Hadj-Salem, H., 1990. Modelling and real-time control of traffic flow on the southern part of boulevard peripherique in paris: Part i: Modelling. Transportation Research Part A: General 24 (5), 345--359.

Payne, H. J., 1971. Models of freeway traffic and control. Mathematical models of public systems.

Punzo, V., Borzacchiello, M. T., Ciuffo, B., 2011. On the assessment of vehicle trajectory data accuracy and application to the next generation simulation (ngsim) program data. Transportation Research Part C: Emerging Technologies 19 (6), 1243--1262.

Rajagopal, R., Varaiya, P., 2007. Health of California's Loop Detector Sytem. California PATH Program, Institute of Transportation Studies, University of California at Berkeley.

Rey, A., Jin, W.-L., Ritchie, G. S., 2015. Vehicle trajectory estimation based on newells simplified kinematic wave model with eulerian and lagrangian traffic data. Extended abstract submitted to ISTTT22.

Richards, P. I., 1956. Shock waves on the highway. Operations research 4, 42--51.
Russell, S., Norvig, P., 1995. Artificial intelligence: A modern approach. Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs 25.

Su, H., Zhang, L., Yu, S., 2007. Short-term traffic flow prediction based on incremental support vector regression. In: Natural Computation, 2007. ICNC 2007. Third International Conference on. Vol. 1. IEEE, pp. 640--645.

Sun, C., Ritchie, S. G., Tsai, K., Jayakrishnan, R., 1999. Use of vehicle signature analysis and lexicographic optimization for vehicle reidentification on freeways. Transportation Research Part C 7, 167--185.

Sun, X., Muñoz, L., Horowitz, R., 2003. Highway traffic state estimation using improved mixture Kalman filters for effective ramp metering control. Decision and Control, 2003. ... 1 (1).
URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1272322
Thiemann, C., Treiber, M., Kesting, A., 2008. Estimating acceleration and lanechanging dynamics from next generation simulation trajectory data. Transportation Research Record: Journal of the Transportation Research Board (2088), 90--101.

Turochy, R., Smith, B., 2000. New procedure for detector data screening in traffic management systems. Transportation Research Record: Journal of the Transportation Research Board 1727, 127--131.

USDOT, 2008. Ngsim--next generation simulation.
URL http://www.ngsim.fhwa.dot.gov/
Van Der Voort, M., Dougherty, M., Watson, S., 1996. Combining kohonen maps with arima time series models to forecast traffic flow. Transportation Research Part C: Emerging Technologies 4 (5), 307--318.

Waller, S. T., Kockelman, K. M., Sun, D., Boyles, S., Lin, D.-Y., Ng, M., Seraj, S., Tassabehji, M., Valsaraj, V., Wang, X., 2008. Archiving, sharing, and quantifying reliability of traffic data.

Wang, Y., Malinovskiy, Y., Wu, Y.-J., Lee, U. K., Neeley, M., 2011. Error modeling and analysis for travel time data obtained from bluetooth mac address matching. Department of Civil and Environmental Engineering University of Washington.

Wang, Y., Papageorgiou, M., 2005. Real-time freeway traffic state estimation based on extended kalman filter: a general approach. Transportation Research Part B 39, 141--167.

Welch, G., Bishop, G., 1995. An introduction to the kalman filter.
Williams, B. M., Hoel, L. A., 2003. Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results. Journal of transportation engineering 129 (6), 664--672.

Work, D. B., Tossavainen, O.-P., Blandin, S., Bayen, A. M., Iwuchukwu, T., Tracton, K., 2008. An ensemble kalman filtering approach to highway traffic estimation using gps enabled mobile devices. In: Decision and Control, 2008. CDC 2008. 47th IEEE Conference on. IEEE, pp. 5062--5068.

Yang, H., Gan, Q., Jin, W.-L., 2011. Calibration of a family of car-following models with retarded linear regression methods. In: Transportation Research Board 90th Annual Meeting.


[^0]:    ${ }^{1}$ The cumulative flows are abbreviated as "flows" hereafter

[^1]:    ${ }^{2}$ From an algorithm analysis perspective, to eliminate elements in the first column is $(|N|-1) L$ operations: $|N|-1$ entries need to be multiplied and subtracted, and this needs to be done $|L|$ times). The second would require $(|N|-2)(|L|-1)$ operations. This process has to be repeated $|N|-1$ times.

[^2]:    ${ }^{1}$ In Deng et al. (2013), the entry/exit time of vehicles are assumed to be measured by AVI technology. Then the CMR is equivalent to the market penetration rate of AVI vehicles.

[^3]:    ${ }^{2}$ Note that is Deng et al. (2013) only provided the result of one run.

[^4]:    ${ }^{1}$ The reason to predetermine the free-flow speed is explained in Subsection 5.4.1.

