Chapter 1

Lane Management

Authors: Amir Meshkat and YU Han

Preface

As a part of the TRAIL course on "Advanced Road Traffic Operations" two fundamentally important papers in the field of *lane management* were discussed and this brief review is made accordingly. The main purpose of this is to transfer acquired experience to the fellow PhDs.

1.1 Lane-changing in traffic streams

(Laval and Daganzo, 2006a)

1.1.1 Introduction

Freeway lane-changing has received increased scientific attention during the last decade, but a *quantitative understanding* of its impacts on traffic flow remains elusive. Existing traffic flow models do not address lane-changing phenomena properly, because they treat lane-changing vehicles as a *fluid that can accelerate instantaneously*. This problem is overcome in this paper by modeling each lane as a separate kinematic wave (KW) stream interrupted by lane-changing particles that completely(and temporarily) block traffic, and moves with bounded acceleration. Also one new parameter for *lane-changing demand*, is used to predict the flow transfers between neighboring lanes.

1.1.2 Theoretical Model: The multi-lane kinematic wave module

The continuum formulation

The conservation equation for a single lane, ℓ, in a highway with $n \geq 2$ lanes is:

$$\frac{\partial k_\ell}{\partial t} + \frac{\partial q_\ell}{\partial x} = \Phi_\ell \quad \ell = 1, \ldots, n \tag{1.1}$$

Where the term Φ_ℓ is the net lane-changing rate onto lane ℓ, in units of veh/(time×distance), and related to the one-directional rates by:

$$\Phi_\ell = \sum_{\ell' \neq \ell} \Phi_{\ell'\ell} - \Phi_{\ell\ell'} \tag{1.2}$$

$\Phi_{\ell\ell'}$ is the one-directional lane-changing rate from lane ℓ to lane ℓ'.

- Attention: ℓ' could only be $(\ell - 1)$ and $(\ell + 1)$

Lane-changing rate

The $\Phi_{\ell\ell'}$ must realistically represent the competition between drivers' desires for changing lanes, and the available space capacity in the target lane.

$$(\Phi_{(\ell-1)\ell}, q_\ell, \Phi_{(\ell+1)\ell}) = F(L_{(\ell-1)\ell}, T_\ell, L_{(\ell+1)\ell}, \mu_\ell) \tag{1.3}$$

Where:

- F is a competition mechanism
- $L_{\ell\ell'}$ is a desired lane-changing rate from ℓ to ℓ'
- T_ℓ is a desired through flow on ℓ
- μ_ℓ is the available capacity on lane ℓ

The discrete-time formulation

There are several assumptions here:

- The fundamental diagram (FD) of each lane is triangular with free-flow speed u, wave speed $-w$ and jam density κ.
- All lanes are partitioned into small cells of length Δx and time is discretized into steps of duration Δt
- The following three-dimensional grid is used $(t_j = j\Delta t, x_i = i\Delta x, \ell)$
- Indices i and j will be used to denote the calculated values of a variable at a discrete point (t_j, x_i)

The sending function (demand) for triangular FDs gives the desired aggregate number of advancing moves in Δt:

$$S_{i\ell}^j = \Delta t \min\{uk_{i\ell}^j, Q\} \tag{1.4}$$

The desired number of lane-changing moves in Δt is given by:

$$L_{i\ell\ell'}^j \Delta t \Delta x = \pi_{i\ell\ell'}^j \Delta t S_{i\ell}^j, \quad \forall \ell, \forall \ell' \neq \ell \tag{1.5}$$

Where $\pi_{i\ell\ell'}^j$ is the fraction of choice-makers per unit time wishing to change from lane ℓ to lane ℓ'. Assume this choice probability rate is proportional to the positive speed difference between lanes:

$$\pi_{i\ell\ell'}^j = \frac{\Delta v_{i\ell\ell'}^j}{u\tau}, \quad \forall \ell, \forall \ell' \neq \ell \tag{1.6}$$

τ can be interpreted as the time a driver takes to decide and execute a lane change when the origin lane is stopped and the target lane is freely flowing.

• Comment: A higher speed difference make it harder to change the lanes, that could have also a negative effect on the choice probability rate.

The probability of staying in the same lane in the next Δt will be:

$$1 - \sum_{\ell' \neq \ell} \pi_{i\ell\ell'} \Delta t \tag{1.7}$$

Therefore

$$T_{i\ell}^j \Delta t = (1 - \sum_{\ell' \neq \ell} \pi_{i\ell\ell'}^j \Delta t) S_{i\ell}^j \quad \forall \ell \tag{1.8}$$

And the available capacity is given by the receiving function (supply) for a triangular FD:

$$\mu_{i\ell}^{j}\Delta t = \Delta t \min\{w(\kappa - k_{i\ell}^{j}), Q\} \quad \forall \ell \tag{1.9}$$

When total demand is less than the available capacity all the demands are fulfilled and able to advance to the target cell; Otherwise the incremental transfer (IT) prorates the available capacity to the different origin lanes according to their demands.
It has been shown in (Leclercq, 2004) that if γ_ℓ represents the fraction of the demand able to advance, then the IT result reduces to:

$$\gamma_\ell = \min\{1, \frac{\mu_\ell}{T_\ell + \sum_{\ell \neq \ell'} \Delta x L_{\ell'\ell}}\} \tag{1.10}$$

And our competition mechanism F reduces to :

$$\Phi_{\ell'\ell} = \gamma_\ell L_{\ell'\ell}, \quad \forall \ell' \neq \ell \quad and \quad q_\ell = \gamma_\ell T_\ell \tag{1.11}$$

• Comment: The use of same γ_ℓ for lane changing and through flow?

The constrained motion model

Discrete lane-changing particles: This model ignores that the disruption caused by a lane change depends on the *initial speed* of the lane-changing vehicle and on its *ability to accelerate*. To solve this the basic idea consists in quantizing the lane-changing rates to generate *discrete particles*, and then treating them as *temporary blockages* that move with bounded accelerations. This is possible because the blockages have a *known (zero) passing rate* and trajectories that can be determined endogenously with the *constrained motion (CM) model* of vehicle dynamics.
In the CM model, particles move with maximum acceleration, but are constrained by their own power and the speed of traffic ahead:

$$v(t) = \min\{v_{dwn}(t), v_{des}(t)\} \tag{1.12}$$

Free-motion models for the kinematics of an isolated vehicle capture the physical limitations imposed by roadway geometry on the engine for typical driver behavior.

$$a = a_0(1 - v/v_{max}) - gG \tag{1.13}$$

Where $G(x)$ is the grade at location x expressed as a decimal, v_{max} is the car's maximum speed, and a_0 is the zero-speed acceleration

1.1.3 Empirical tests

General Assumptions

For following numerical experiments assume that:

- $\tau = 3s$
- $u = 96.6km/h$
- $w = -24km/h$
- $\kappa = 93.2vpkp\ell$

The car type used in these numerical demonstrations is a high performance car, defined with:

- $v_{max} = 155km/h$
- $a_0 = 4.3m/s^2$
- and $g = 9.81m/s^2$

Lane-drops

Recent experiments (Bertini and Leal, 2006) show a consistent reduction in discharge flow after the onset of congestion at bottlenecks caused by lane-drops. This suggests that the drop in discharge rate may be caused by lane changers, since approaching mainline vehicles are forced away from the shoulder lane.
The proposed model was applied to a $0.5\,km$ 3-lane freeway with a lane-drop at $x = 0.33\,km$. At $t = 0$ the freeway is empty and the input demand is held constant at $1242\,vph$ on the median and middle lanes (lanes 1 and 2, respectively) and $416\,vph$ on the shoulder lane (lane 3).
The cumulative count curves across all lanes, $N(t)$, produced by the proposed method. Curves are plotted on an oblique coordinate system with a *background flow* of $2630\,vph$. They show that flow drops from $2900\,vph$ to $2630\,vph$ around $t = 10\,min$.This is a 9.3% drop.

Moving obstructions

(Munoz and Daganzo, 2002) describes the results of several experiments where obstructions with a range of controlled speeds v were introduced in 2-lane and 3-lane freeways.
It was found that there was a reproducible relation between the capacity of the moving bottleneck, Q_m, and v:

$$Q_m = Q_m(v), \quad with \quad \frac{dQ_m}{dv} > 0 \quad for \quad v > 50km/h \tag{1.14}$$

To test the proposed theory the experiments in (Munoz and Daganzo, 2002) were simulated for a broader set of conditions:

$$v \in \{0, \ldots, 80\}\, km/h \quad \textbf{and}, \quad n \in \{2, 3, 4\} \tag{1.15}$$

The (normalized) capacity, $\rho = Q_m(v)/Q_m(u)$, versus v

The model predicts two regimes. Regime 1 ($v > 20km/h$) where $\rho(v)$ increases, and regime 2 ($v < 20km/h$) where $\rho(v)$ slightly decreases. The two-regime phenomenon is qualitatively explained by the spatial distribution of the lane changes.
The k-maps of one simulation for two bottleneck speeds, $v = 1$ and $32km/h$ are shown here. Lane-changing locations are depicted as bold red dots.
Lane-speed differences are greater close to the bottleneck. The reason for the capacity trend in regime 1 is a direct consequence of this effect, because as v increases, lane-speed differences drop and lane-changing becomes less and less disruptive. There is a critical distance from the bottleneck beyond which lane-changing has no effect. Furthermore, for very low speeds a significant number of lane changes occur beyond this critical distance. The trend of regime 2 is then explained, because as $v \to 0$ more and more lane changes take place far upstream from the bottleneck where lane-speeds are similar.

1.1.4 Discussion

A four-parameter multi-lane hybrid model for traffic flow that recognizes the bounded accelerations of lane-changing vehicles has been introduced. The model appears to explain the reduction in flow observed after the onset of congestion at freeway *lane-drops* and the relationship between the speed of *moving bottlenecks* and their capacities. Traffic managers may be able to increase capacity by forbidding lane changes and/or posting speed advisories at key locations upstream of bottlenecks.

1.2 Freeway Traffic Oscillations and Vehicle Lane-Change Maneuvers

(Ahn and Cassidy, 2007)

1.2.1 Background

The author introduced previous research about oscillations. It was described as a car following behavior. Drivers try to maintain their desired spacing and speed, the inability to perceive speeds and spacing lead to instabilities, which is the description of oscillation. However, these model predictions do not always match real measurements. For example, the observations of real freeway indicate the acceleration and deceleration period of oscillations last several minutes. Car following models produce instabilities with periods on the order of a driver reaction time. In addition, the amplitude growing of oscillation was largely neglected in the car-following model. It was just assumed to happen at the vicinity of busy ramps. A study from (Treiterer and Myers, 1974) found that densities within the platoons remained high as they underwent accelerations. These findings are used as the support for the car-following models. The author emphasized lane changing should be the real reason, because during acceleration, drivers from neighboring lane try to insert into platoons, which leads to a high density. Some oscillations have been observed in the platoons of tunnels, where lane changing was prohibited. In these cases, oscillations might be explained by car-following models. However, in this paper, the author provided compelling evidence that lane changing is the real reason of freeway oscillation.

1.2.2 Data

A 6-km-long stretch of San Francisco Bay Area was considered as the test case of this paper. There are downstream bottlenecks in both direction of this stretch. Vehicle counts, occupancies, time-mean speeds were collected for macroscopic analysis; video surveillance data were collected for microscopic analysis.

1.2.3 Macro-level analysis

The author provided some clues that the formation and growth of oscillations are relevant to lane changing. A lane specific time series curve of vehicle speed is shown. The difference between the cumulative value of speed and time weighted average speed was plotted, to show the speed deviation, which are oscillations. The oscillation amplitude of the upstream most detector is much higher than that of the downstream most detector, which indicates that the oscillation grows when they propagates upstream. This trend was confirmed by a Root Mean Squared Error curve. The curve shows that the downstream most detectors have smallest RMSE value. Lane changing event is relevant to the oscillation was confirmed by a coefficient plot of speed deviation. The downstream most detector show a relatively low speed coefficient, which implies oscillation emerged separately in each lane. Figures from east bounds show similar results. The coefficient of the downstream most detector of the east bound is higher than that of the westbound, because the bottleneck of the eastbound is farer than that of the westbound, so the oscillation of the eastbound had already synchronized.

1.2.4 Micro-level analysis

Vehicle speeds were individually extracted from video and the vehicle trajectories were constructed. Two 100-meter-long segments, which locate near the downstream bottleneck, were chosen to analyse the vehicle trajectory. 46 vehicles were tracked, and their moving average speed on two segment were plotted separately. This figure revealed the formation of an oscillation. The oscillation started at the upstream segment, with the great speed reduction of vehicle 11. Vehicle trajectory unveiled its cause. An additional vehicle changed its lane and inserted itself in front of vehicle 11. From the trajectories the author also show that the deceleration amplified for further vehicles, which indicated that car-following behavior can contribute to the oscillation. According to the analysis of more than 1470 vehicle speeds, two criteria were judged the requirement of the formation of oscillation. The first criterion is that the greatest difference in the moving average vehicle speeds on a segment should be larger than 7km/h. The second criterion is that a criterion 1 was not measured on the downstream segment, to exclude that the oscillation was not propagated from the downstream segment. Using these criteria, ten instances of formation were detected. Trajectories show that lane changing were always the triggering events.

1.2.5 Growth

The author analyzed the oscillation growth by studying the eastbound traffic. The speeds of 51 vehicles over two continuous segments were shown. Both two segments experienced the oscillation, but the speed drop of upstream segment is higher than the downstream segment, which indicates the oscillation was amplified. Speed of vehicle 29 was the first one, which was lower than any observed speed of the upstream segment. The trajectories unveiled that this oscillation growth is caused by a vehicle, which inserted in front of vehicle 29. Another insertion vehicle, vehicle 35, did not amplify the oscillation, but prolonged the period over which vehicle traveled at a lower speed. According to the measurement of 1500 speeds, the author proposed two criteria to judge the occurrence of oscillation growth. The first one is the same as the criteria of judging the formation of oscillation. It is not difficult to understand because if you want to judge whether the oscillation propagates, you should first makes sure it is an oscillation. The second criterion is that the minimum speed difference between two segments had to be at least 1 km/h. Data shows that growth was more likely to be generated by lane changes made into spacing that were small. The spacing of triggered growth has a mean value 34m, while other spacing 54m.

1.2.6 Conclusion

In this paper, the author unveiled a causal relation between vehicle lane-changing and traffic oscillation. From vehicle trajectory data, the author shows how lane changing triggered the formation of oscillation and how it contributes to the growth. From the large amount of data, they did not find the spontaneously formed oscillations. This paper subverted previous studies that the oscillation is caused by a car-following behavior. Some puzzles still existed, for example, why lane changing caused longer oscillation period, why the synchronization enlarged after oscillation propagated.

1.3 Spatiotemporal Effects of Segregating Different Vehicle Classes on Separate Lanes

(Cassidy et al., 2009)

1.3.1 Introduction

The paper explores some of the impacts of setting aside road lanes for the exclusive use of select vehicle classes, and it makes three contributions. The first is methodological: it illustrates the importance of analyzing freeway data in full spatiotemporal detail. The second is physical: data reveal that carpool lanes are not as damaging as previously reported. In fact, these lanes are found to smooth traffic in adjacent lanes so much (by diminishing disruptive vehicle interactions near bottlenecks) that even substantially underutilized carpool lanes can increase bottleneck discharge flows. The third contribution is theoretical: it uses the smoothing phenomenon to show how the judicious deployment of bus-only lanes on freeways and city streets can favorably affect not just buses, but also cars.

1.3.2 Methodological Contribution: Spatiotemporal Analysis

The EATO study used time series of vehicle speeds and flows (by lane) as metrics to assess the carpool lanes' impacts on freeway traffic. Like in many other studies, these data were measured at single detector stations, and were not analyzed together with data from neighboring detectors. EATO identified periods when queues persisted atop a detector, and correlated these with carpool-lane operating times. But this methodology of examining only data from single detector stations can not identify the locations where these queues initially formed.
The EATO study claims without examining demand effects that queues arose in these General Purpose (GP) lanes because they were eventually in short supply; i.e., demand among Low Occupancy Vehicles (LOVs) presumably grew while the median lane was unavailable for their use, and this supposedly pushed the GP lanes into the congested regime. However, the single-detector station methodology is inconclusive, and the events should reassess in spatiotemporal detail.

1.3.3 New Physical Finding: The Smoothing Effect

The video data unveil the car- pool-lane effect on both the bottleneck's formation, and its discharge flow.

Bottleneck Formation

Vehicle arrival times were manually extracted from the videos and, as is customary, cumulative curves of vehicle count were plotted on an oblique coordinate system (O-curves). The video data establish that the flow reduction was triggered by a queue that first formed in the shoulder lane due to pulses of merging vehicles, and then spread to all lanes; the carpool lane had no role in this. Moreover, the lane did not impede bottleneck flow and prolong congestion, despite being underutilized.

Subsequent Effect of the Carpool Lane

The video data reveal that drivers (of LOVs) began avoiding the median lane shortly before the carpool restriction went into effect; and that these driver responses began after the capacity drop had already occurred. The carpool-lane flow eventually dropped, and yet the total flow across all lanes (including the carpool lane) remained quite steady. The bottleneck's total discharge flow returned to its highest rate, when carpool-lane flow was lowest. This indicates that the diminished carpool-lane flow was compensated by increased queue discharge flow (capacity) in the adjacent GP lanes.

1.3.4 Theoretical Application: The Effect of Bus Lanes

This section shows how the deployment of bus-only lanes can in some cases improve the flow of cars. Because the magnitude of the smoothing effect for roadways with cars and buses is not yet predictable, its system-wide impacts was parametrically predicted. We will do this for rotationally symmetric closed-loop beltways with access and egress via on- and off-ramps because as explained in (Daganzo and Cassidy, 2006), this is the least favorable environment for a special-use lane. Suppose that a beltway cannot be metered very restrictively, so that queues form on it. Even in this case, segregation can increase the beltway's flow (and therefore its input and output flows). This is a good thing because delays and queues of cars would then diminish outside the beltway (e.g. on its on-ramps and connecting streets), without increasing on the beltway itself; while bus users would also benefit by enjoying higher travel speeds.

1.3.5 Conclusion

There are many possible causes of roadway traffic congestion, including accidents, roadwork activity, high merge demands and special-use lanes; and one needs to rule-out all other possibilities before attributing congestion to any one cause. This paper has shown that analysis of time series data alone, without also considering a system's spatial component, will not provide a complete picture of how special-use lanes affect traffic, and can produce misleading results.

Chapter 2

Summary Advanced Road Operations - Development and use of Variational Theory in Traffic Flow Modelling

Kai Yuan & Dorine C. Duives

2.1 INTRODUCTION

The following pages summarize the development and use of variational theory by means of the following key studies:

- Newell, G.F., (1993). A simplified theory of kinematic waves in highway traffic, Part 1: general theory. Transportation Research Part B, 4, 281-287.
- Daganzo, C., (2005). A variational formulation of kinematic waves: basic theory and complex boundary conditions. Transporation Research Part B, 39, 187-196.
- Leclercq, L., Laval, J., Chevallier, E., (2007). The lagrangian coordinates and what it means for first order traffic flow models. In:. Transportation and Traffic Theory 2007. Allsop, R.E., Bell, M.G.H., Heydecker, B.G., (Eds.). Elsevier. 735 753.
- Leclercq, L., (2007). Hybrid approaches to the solutions of the Lighthill-Whitham-Richards model. Transportation Research Part B, 41, 701-709.
- Daganzo, C., Geroliminis, N., (2008). An analytical approximation for the macroscopic fundamental diagram of urban traffic. Transportation Research Part B, 42, 771-781.

Each paper will be discussed separately, while retaining the original structure and notation of the papers as much as possible. Therefore, there might be a difference in notation or definitions between the papers. These differences have been mentioned explicityly. The papers are discussed in the chronological order mentioned above. Since it is a summary, it has been attempted to detail the development of the papers, and not the intricate details. For these specific details and a complete overview of the mathematical development one is referred to the respective papers.

2.1.1 Intro to Variational theory

Figure 2.1 shows an overview of macroscopic modeling process. It consists of three steps: establishing kinematic wave model, discretizing the model, solving models for solutions. The kinematic wave model is a combination of vehicles conservation rule and fundamental diagram. When applying the kinematic wave model, researchers firstly discretized it. For example, the space and time can be divided into grid

cells and time steps respectively in Eulerian coordinate, or the vehicles and time are divided into groups and time steps respectively in Lagrangian coordinate. Finally numerical methodologies are used to solve the discredized model. The variational theory is one of the numerical methodologies. Researchers apply variational theory to update parameters, e.g. $N(x,t)$, in the discredized model.

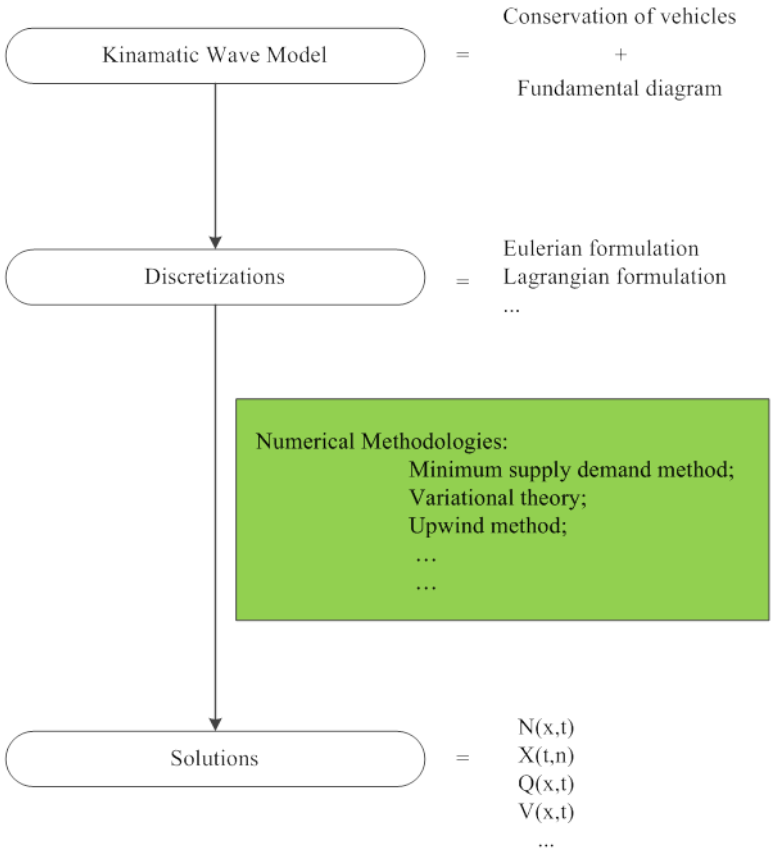

Figure 2.1: Process of solving the LWR-model

2.2 NEWEL 1993 - A SIMPLIFIED THEORY OF KINEMATIC WAVES IN HIGHWAY TRAFFIC

2.2.1 Background

Lighthill and Whitham (1995) and Richards (1956) independently proposed a theory of one-dimensional wave motion which could be applied to highway traffic flow. They postulate that there exists a functional relation between the flow q and the density k. In their work, two things are assumed:

- Road conditions are homogeneous, they might vary in space but not in time.

$$k(x,t) = k^*(q(x,t),x) \tag{2.1}$$

- There are no entering nor exiting vehicles

$$\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0 \tag{2.2}$$

This gives rise to the partial differential equation:

$$\frac{dk^*}{dq}\frac{\partial q}{\partial t} + \frac{\partial q}{\partial x} = w\frac{\partial q}{\partial t} + \frac{\partial q}{\partial x} = 0 \tag{2.3}$$

where $1/w$ represents the wave speed. If assuming that the traffic state q is constant along a characteristic curve with the wave speed, then one can solve the traffic state everywhere in the system if the initial/boundary conditions are known. These characteristic curves may intersect, yielding multi-valued flows. A solution is the use of shockwave theory with a separate equation for the shock path. Numerical solutions are typically rather tedious.

2.2.2 Cumulative curves

Newel presents the use of cumulative curves to compute the traffic state everywhere on an (x, t)surface,were $A(x, t)$ is cumulative number of vehicles to pass some location x by time t starting from the passage of some reference vehicle.

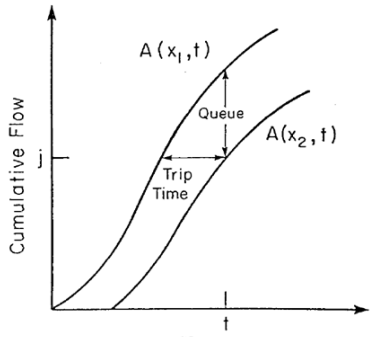

Figure 2.2: Cumulative flow at two locations Newell (1993)

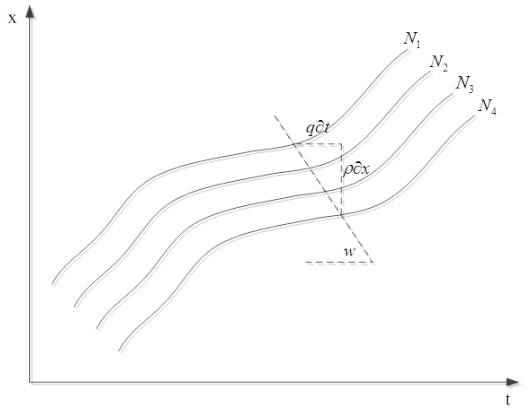

Figure 2.3: Vehicle trajectories and one characteristic curve

If onw draws the curves $A(x_1, t)$ and $A(x_2, t)$ for two locations $x_!$ and x_2 on the same graph (see figure 2.2), the vertical distance between the curves at time t represents the number of vehicles between x_1 and x_2, and the horizontal distance between the curves at height j the trip time of the jth vehicle (Given that vehicles do not pass each other). As shown in figure 2.3, there is a moving observer which moves in an oppsite direction with a speed of w .w is negative in figure 2.3. The trajectory of the moving observer is referred to as characteristic curve in this chapter, shown as a dashed line in figure 2.3. Newell deduces that the change in A can be determined based on the knowledge of q and k along the characteristic curve with wave speed w from (2.1):

$$dA = -kdx + qdt = (-k + qw)dx \tag{2.4}$$

Thus, if one knows $A(x_0, t_0)$ and $q(x_0, t_0)$ at some boundary point, one can determine $A(x, t)$ at all points along the characteristic curve through (x_0, t_0). If we accordingly interpret $A(x, t)$ as a surface in 3-dimensional space, the surface $A(x, t)$ is a "ruled surface". That is to say, a surface generated by a familty of straight lines. From appropriate boundary data one could quite easily construct a surface and determine $A(x, t)$ everywhere.

2.2.3 Shock conditions

When characteristic curves intersect, the surface $A(x, t)$ gives a multiple-valued function of x and t. But since, the number of vehicles which can pass any point in time should not exceed any "constraint" generated by the motion of vehicles at earlier times, Newel claims that the solution is the lower envelope of the multiple-valued solution derived from properly set initial or boundary conditions. As a result:

- A is continuous, eventhough derivatives q and k might not be.
- The shock path automatically emerges as a discontinuity in the solution (it is handled implicitly).
- solution also works if x or t is discrete.

2.2.4 Applicability

The use of cumulative curves to estimate the traffic state of a x, t-surface can be usefull for a variety of problems:

- Analysis of highway sections between on-/off-ramps
- Analysis of traffic flow near traffic signals
- Analysis of trajectory of a vehicle from solution

Where the initial conditions for A can be distillated from the initial density and the boundary conditions from the boundary flow.

2.3 DAGANZO 2005 - A VARIATIONAL FORMULATION OF KINEMATIC WAVES: BASIC THEORY AND COMPLEX BOUNDARY CONDITIONS

2.3.1 Introduction

The paper of Daganzo (2005b) embroiders on the work of Newell (1993). It starts of by noting the cumulative curve $N(x, t)$ at time t and location x should satisfy the following partial differential equation:

$$N_t = Q(-N_x, t, x) \rightarrow q = Q(k, t, x) \tag{2.5}$$

where the subscripts note the partial derivatives of N. According to Daganzo, Newell (1993) and Luke (1972) proposed that if N has been defined on a boundary D, than the value of N at a point P is given by the rule:

$$N_P = min\,(B_W + \Delta_W) \tag{2.6}$$

were W notes all space-time paths, $x(t)$, directed in the direction of increasing time with specific beginning and ending times within the set of all paths from D to P and within the set of all directed wave paths emanating from D. B_w represents the start-up costs at boundary D (value of N), while Δ_W represents the traveling costs along the characteritic curve (increase/decrease of N).

A solution to 2.5 is a surface in 3-dimensional space formed by a family of characteristicss which satiesfy the following system of ordinary differential equations:

$$x' = Q_k \tag{2.7}$$

$$k' = -Q_x \tag{2.8}$$

$$N' = q - kQ_k \tag{2.9}$$

$$q' = Q_t \tag{2.10}$$

Daganzo shows that if this problem is well-posed, a unique solution to 2.1 exists. A problem is well-posed if the following things hold:

- There is a unique characteristic associated with each boundary point D.
- The running costs (Δ_w) only depend on the values of x, t and k at the root of the wave.
- For every space-time point P in the solution space a characteristic wave can be found which passes through this point.

However, finding a solution by means of the methods proposed by Newell (1993) or Luke (1972) still cumbersome for general problems due to the difficulties in the identification of relevant set of paths.

2.3.2 Variational principle

Consider a wave that yields the minimum solution (2.6) at point P. In that case, the rate at which cars are overtaken by an observer moving with the wave can be defined as follows:

$$r = Q(k, t, x) - ku \tag{2.11}$$

where r is the overtaking rate and u the wave speed. In Figure 2.4, ΔA is shown as line 3. Since the fundamental diagram is always concave, so we can update Equation 2.11 into Equation 2.12. Since the fundamental diagram is concave, k can be eliminated, rendering:

$$R(u, t, x) = r = \sup_k \{Q(k, t, x) - ku\} \tag{2.12}$$

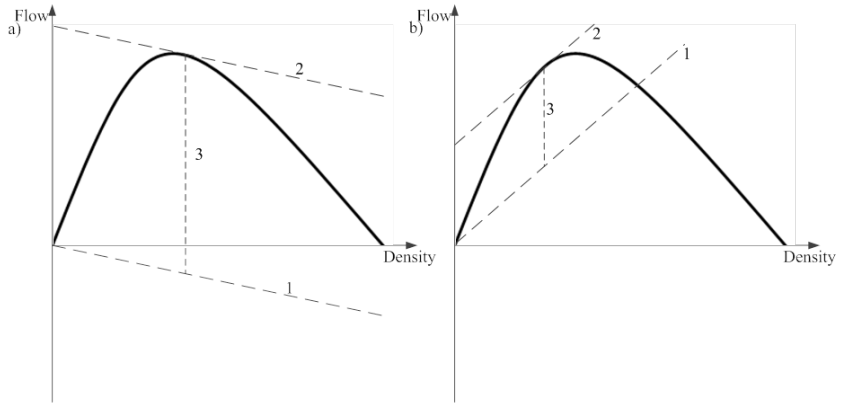

Figure 2.4: Fundamental diagrams. The moving observer is moving backward (a) and forward (b). Line 1 indicates the parth of moving observer. Line 2 indicates the corresponding characteristic curve. Line 1 is parallel to Line 2. Line 3 indicates ΔA.

The objective function is maximized by k^* if and only if k^* is such that $Q_k = u$. That is to say, one is traveling with the wave speed. Since the result implies that the rate at which N changes along a wave is an intrinsic property of Q, and does not depend on the boundary data or the solution, Δ_W can be written as the integral over time of a function $R(x', t', x)$, which is independent of the input data. Furthermore, the correct general formula for any path $\mathscr{P}$ going from B to P is the integral of the rate at which N changes along the apth, this is:

$$N_P - N_B = \int_{t_B}^{t_P} |Q(k, t, x) - kx'| \tag{2.13}$$

where $x(t)$ is a trajectory of $\mathscr{P}$. Because a continuous solution of 2.5 satisfies $N_P - N_B <= \Delta(\mathscr{P})$, the functional $R(u,t,x)$ can be applied to any "valid" path, and not just waves. As a consequence, (2.6) can be replaced by:

$$N_P = min\left(B_W + \Delta(\mathscr{P})\right) \tag{2.14}$$

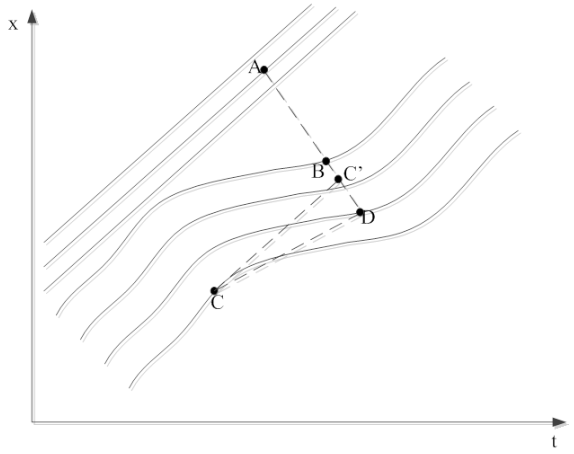

Figure 2.5: Vehicles trajectories and three characteristic curves.

In Figure 2.5, the difference between the $N^{A-D}(x_D,t_D)$ and $N^{B-D}(x_D,t_D)$ is due to some external reasons which influced the fourth drivers behavior. That is, without the external impacts, the deduced $N(x_D,t_D)$ should be a fixed constant, that is $N^{B-D}(x_D,t_D) = N^{C-D}(x_D,t_D)$.In another words, the vehicle number at one point are indepent of the starting boundary point and path.

2.3.3 practical issues

Daganzo notes three practical issues, which are further elaborated upon in Daganzo and Menendez (2005) and Daganzo (2005c).

- *well-posedness:* The problem needs to be well-posed, otherwise detectable negative flows might occur in the solution. As such, time-dependent fundamental diagrams might be problematic.
- *solution networks:* One can approximate the set of paths V with a discrete graph of edges in space-time, where the slopes of the edges cover all wave speeds and the costs of the edges are $\Delta(\mathscr{P})$. The minimum problem can accordingly be formulated as a shortest path problem from some boundary point B to the solution point P.
- *Complex boundary conditions:* Adding shortcuts to the solution network can reduce N. This idea can be used to add standing or moving bottlenecks inside the modelled space-time area. This changes the solution but not the solution method. As a result, the properties of the method (e.g. single-valuedness, continuity and stability) are preserved.

2.4 LECLERQ 2007 - THE LAGRANGIAN COORDINATES AND WHAT IT MEANS FOR FIRST ORDER TRAFFIC FLOW MODELS

2.4.1 Introduction

The paper of Leclercq et al. (2007) embroiders on the work by Newell (1993) and Daganzo (2005b). Newell proposed the use of the cumulative curve and conjectured that the LWR solution is the lower envelope of the multiple values of $N(x,t)$. Daganzo used variational theory to prove this conjecture, which reduced the LWR model to the solution of the Hamilton-Jacobi equation $\partial_t N = Q(-\partial_x N)$. This

new approach opened the door to powerful numerical methods based on shortest-path algoritms in (x,t) coordinates. Leclercq et al. (2007) takes it one step further, and formulates the LWR model in the transformed coordinate system (N,t).

2.4.2 The LWR model in lagrangian coordinates

The conservation law in Lagrangian coordinates was first introduced by Courant and Friedrich in 1948 in case of gas dynamics. For traffic flow 2.1 becomes:

$$\partial_t s + \partial_n V^*(s) = 0 \tag{2.15}$$

where s corresponds to the spacing $(1/k)$ and $V^*(s)$ to the concave fundamental relation between velocity and spacing. In order to solve this problem, the following partial differential equation has to be solved:

$$\partial_t X = V^*(\partial_n X) \tag{2.16}$$

Leclercq et al. (2007) proves that $\partial_t X$ and $V^*(\partial_n X)$ both exist. Therefore, it is possible to formulate the Lagrangian variational principle similarly to the Eulerian version of the problem. As a consequence, all results of Daganzo (2005b) can also be applied to the Lagrangian formulation of the LWR model. That is to say, the value of X at point P in the (N,t) plane, can be expressed as follows:

$$X_P = min\left(B_P + \Delta(\mathscr{P})\right) \tag{2.17}$$

The cost rate r on a wave path is given by $d_t X$, which represents the speed of the Eulerian characteristic associated to the passing rate u. Since $V^*(s)$ is concave, one can express r as a function fo the adopted velocity, the spacing and the wave speed (see eq. 2.18).

$$r = R(u) = \sup_s \{V^*(s) - su\} \tag{2.18}$$

The cost on a Lagrangian valid path $\mathscr{P}$ thus becomes:

$$\Delta(\mathscr{P}) = \int_B^P R(n'(t))dt \tag{2.19}$$

Which computes the value of s which maximizes the distance between the speed/spacing relationship $(V^*(s))$ and the passing rate r.

2.4.3 numerical resolution

Given that $V^*(s)$ is non-decreasing, the Godunov scheme, which provides the solution to the problem, reduces to an upwind method:

$$s_i^{t+\Delta t} = s_i^t + \frac{\Delta t}{\Delta n}\left(V^*\left(s_i^t\right) - V^*\left(s_{i-1}^t\right)\right) \tag{2.20}$$

That is to say, the spacing s of vehicle i in timestep $t+\Delta t$ depends on the the spacing s of vehicle i in timestep t plus the distance which either gained or lossed during a short timeinterval due to the velocity differential between vehicle i and the previous vehicle $i-1$. When supposing that Q is triangular, the CFL condition is satisfied as an equality and some simplification, the Lagrangian Godunov scheme can be expressed in $X(n,t)$ as follows:

$$X(n,t+\Delta t) = min\left(X(n,t) + v_m\Delta t, X(n-\Delta n,t) - w\Delta t\right) \tag{2.21}$$

which can be understood as the minimum of the position which vehicle n at time t would reach while moving along with free-flow speed v_m for a timeinterval Δt, and the position that the queue could reach in a similar time-interval if the upstream vehicle $n-1$ would be in congestion. Leclercq et al. (2007) shows that this scheme provides the exact solution.

2.4.4 Numerical errors of the godunov scheme in eulerian coordinates

The paper also obtains insights about the nature of teh numerical viscosity introduced by the Godunov scheme in Eulerian coordinates. A comparison (see figure 2.6) shows that while the solution to in Lagrangian coordinates is exact, numerical viscosity arises for shockwaves propagating backwards. With the exception of isosceles FDs, numerical errors always arise when N_B is the minimum in the variational form of the Godunov scheme for Eulerian coordinates. This happens for shockwaves propagating backwards. Shockwaves propagating forwards do not induce numerical errors when the CFL condition is satisfies as an equality. Leclercq et al. (2007) quantified the upper-bound of the errors in the Eulerian Godunov scheme for shockwaves to be:

$$e_j = (1-\alpha)^{\frac{1}{\alpha}} (l-h)\,\Delta x \tag{2.22}$$

where l and h represent to neighbouring cells with a low and high density. As such, the error is proportional to Δx and to the density difference. Since the error is also decreasing in α, the more the congested brach of the FD is flat, the more important the error in the scheme are becoming.

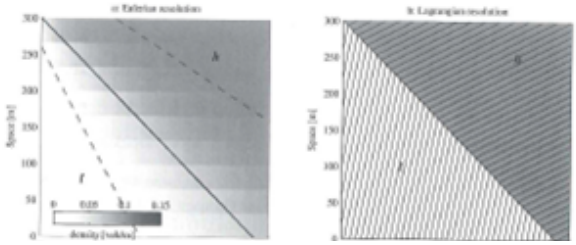

Figure 2.6: Cumulative flow at two locations Newell (1993)

2.4.5 Extensions using the lagrangian variational principle

The paper mentions a number of extensions of the basic theory, i.e. a moving bottleneck, and distinct vehicle characteristics. Below the changes with respect to the basic theory are mentioned.

Moving bottleneck

In case of a exogenous moving bottleneck can be represented as a moving boundary conditions. These moving conditions represent a shortcuts in the solution network $X(n,t)$. Daganzo (2005c) detail this approach. The cost rate of each arc of the speed of the bottleneck v_b, while the slope of the arc in (n,t) is the bottleneck's passing rate r_b.

The paper proposes the following general numerical solution method:

1. Include the shortcut arc in the solution network
2. Calculate the value of X.

$$CaseA \rightarrow B_j = min\left(B_{j-1} + v_b\Delta t, X\left(n,t\right) + v_m\Delta t\right) \tag{2.23}$$

$$CaseB \rightarrow B_j = min\left(B_{j-1} + v_b\Delta t, X\left(n-\Delta n,t\right) - w\Delta t\right) \tag{2.24}$$

3. If the optimum path to B_j does not include the shortcut arc, then set r_b such that the value of X at B_j is the same on both paths and recalculate the value of X at the grid-point by applying to possible paths (case A), and (case B).

Vehicle characteristics

The solution network in Lagrangian coordinates is a natural environment for introducint vehicle characteristics, since the coordinate system moves with the flow. The vehicle-specific characteristic do not modify the lay-out of the solution network, only the characteristics of the arcs. Examples of these changes are:

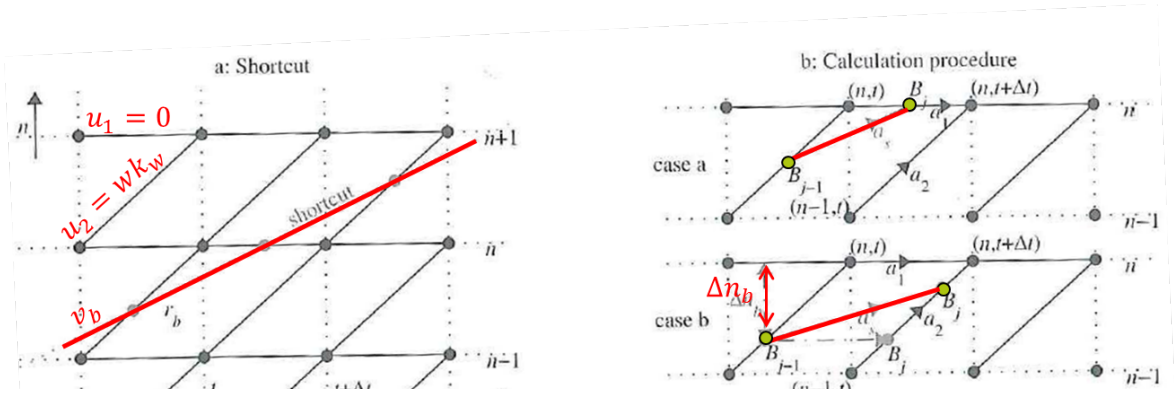

Figure 2.7: Solution network including shortcut which represents moving bottleneck

- *Different free-flow speeds*: change the cost rate of the horizontal arcs of the solution netwerk to represent the vehicle-specific speed v_m^n.
- *Different reaction times*: change the slope w of the slanted arcs between two consecutive vehicles varies.
- *Bounded vehicle acceleration*: change the cost rate to change over time according to $v_m^n(t + \Delta t) = v_m^n(t) + \alpha(t)\,\Delta t$.

Leclercq et al. (2007) provides a short table of the possibilities to implement the complexity by approach (Lagrangian vs. Eulerian), see the table 2.1. Furthermore a discussion of the exactness of the solution by approach with respect to the implemented fundamental diagram, see table

Table 2.1: Extension implementation complexity by approach (++:very easy; + easy; - not so easy; – difficult; ? not known as yet)

	Lagrangian approach		Eulerian approach	
	Godunov scheme	Variational method	Godunov scheme	Variational method
Q triangular + rectangular lattice	√	√	×	√ only if v_w/w is an integer and memory is used
Q triangular	√	√	×	√ only for geometric networks
Q PWL	×	√	×	×

Table 2.2: Exactness by approach (X: not exact)

	Lagrangian approach		Eulerian approach	
	Godunov scheme	Variational method	Godunov scheme	Variational method
Q triangular + rectangular lattice	√	√	×	√ only if v_w/w is an integer and memory is used
Q triangular	√	√	×	√ only for geometric networks
Q PWL	×	√	×	×

2.5 LECLERCQ 2007B - HYBRID APPROACHES TO THE SOLUTIONS OF THE "LIGHTHILL-WHITHAM-RICHARDS" MODEL

2.5.1 Introduction

Traffic flow models can be described either microscopically or macroscopically. Hybrid models are able to combine both descriptions, which makes it possible to treat different parts of the network with different levels of detail. Leclercq (2007) proposes a method to treat the interface between a macroscopic continuum model in Eulerian coordinates and a microscopic model which uses the Lagrangian coordinates. According to Leclercq previous works did consider that at the interface between the models, both models coexist. This does however propose some limitations, being:

- virtual vehicles are necessary which obey fuzzy rules
- the proposed models are constrained to have the same description for all entering and outgoing links

According to Leclercq the approach proposed in this paper overcomes these shortcomings.

2.5.2 Macroscopic resolution of the LWR

The continuum formulation of the LWR in Eulerian (x, t) coordinates is:

$$\partial_t K + \partial_x (KV) = 0 \tag{2.25}$$

When solving this equation by means of a Godunov scheme, the following update rule for the density is found:

$$K_i^{t+\Delta t} = K_i^t + \frac{\Delta t}{\Delta x} \left(Q_{i-1}^{t \to t+\Delta t} - Q_i^{t \to t+\Delta t} \right) \tag{2.26}$$

The scheme does require the Courant-Friedrichs-Lewy (CFL) stability condition $\Delta t <= \frac{\Delta x}{V_f}$ to be satisfied.

2.5.3 Microscopic resolution of the LWR

The microscopic formulation of the LWR in Lagrangian (n, t) coordinates boils down to:

$$\partial_t s + \partial_N V_e^* (s) = 0 \tag{2.27}$$

The solution by means of a Godunov scheme, in this case becomes:

$$d_i^{t+\Delta t} = s_i^t + \Delta t \left(V_e^* \left(s_{i-1}^t \right) - \left(V_e^* \left(s_{i-1}^t \right) \right) \right) \tag{2.28}$$

Also in this case the CFL condition $\Delta N >= \max_s \left| \partial_s \left(V_e^* (s) \right) \right| \Delta t$ needs to be satisfied.

2.5.4 Coupling resolutions

To stich the macroscopic and the microscopic descriptions, interfaces translating boundary conditions are defined. Leclercq uses resevoirs that stock vehicles when arriving at the boundary and spread the right value of flow. For an exact description of the procedure one is referred to Leclercq (2007). He formulates both interfaces separately:

micro2macro interface

The demand during a period Δt is expressed as follows:

$$\Delta \left(x_0^-, t + \Delta t \right) = \begin{cases} V_e^* \left(s_l^t \right), & \textbf{if } s_l^t >= s_c \\ Q_c, & \textbf{if } s_l^t < s_c \end{cases} \tag{2.29}$$

where s_c is the critical spacing. As such, if the spacing of vehicles just behind the boundary at the moment a vehicle crosses the boundary is less than the critical spacing, the demand is capacity. When the spacing is more than the critical spacing, the fundamental diagram $v - s$ determines the spacing and as such the demand. The exit flow Q_c can accordingly be calculated as the minimum between m link demand (2.29) and the M link supply $\Omega\left(K_{0+}^t\right)$. In practice, a resevoir is managing the conservation condition at x_0. The capacity of the resevoir is 1. Over time the resevoir empties with the exit flow rate. Accordingly, every time a vehicle wants to enter, the previous vehicle should have been distributed entirely (resevoir = 0). When the new vehicle enters the resevoir is becomes 1 again.

Macro2micro interface

When transport occurs from the macroscopic to the microscopic model, the supply should be transformed. Leclercq provides the following formula:

$$\Omega\left(x_0^+, t \to t + \Delta t\right) = \begin{cases} Q_c & \text{if } s_n^t >= s_c \\ \frac{V_e^*\left(s_n^t\right)}{s_n^t} & \text{if } s_n^t < s_c \end{cases} \tag{2.30}$$

In this case the supply is computed based on the spacing of the vehicles which are just in front of the boundary. The input flow is calculated as the minimum between the microscopic link supply (2.30) and the macroscopic link demand $\Delta\left(K_{0-}^t\right)$. In practice, also at this interface a resevoir manages the flow. This resevoir has a begin value of 0 and fills up over time. At the end of each timestep in which the resevoir is full (resevoir = 1), a vehicle is released into the microscopic model.

2.5.5 Simulation results

The approach has been tested using four typical flow cases which might arise at a highway, being:

1. Case A: supply variation just downstream of the micro2macro interface
2. Case B: a shockwave coming from the Macroscopic link
3. Case C: demand variations just upstream of the macro2micro interface
4. Case D: a shockwave coming from the microscopic link

The results show that the wave's propagation at both interface levels is correct. The first case (A) shows that the exit flow fits the supply variation without delay. The third case (C) suggests that the input flow fits the analytical solution of these demand variations. While cases B and D show that a shockwave can properly travel through the interface.

2.6 DAGANZO AND GEROLIMINIS 2007 - AN ANALYTICAL APPROXIMATION FOR THE MACROSCOPIC FUNDAMENTAL DIAGRAM OF URBAN TRAFFIC

2.6.1 Introduction

Daganzo (2005a) and Daganzo (2007a) have proposed that traffic can be modelled dynamically in large urban regions at an aggregate level if such regions exhibit two features: a reproducible "macroscopic fundamental diagram" (MFD) and a robust relation between the neighborhood's average flow and its total outflow. Invariance is important because knowledge of a neighborhood's MFD allows decision-makers to use verifiable and robust demand-side policies to improve mobility. Developing a universal recipe that can be used for signal-controlled networks, however, is a challenging task because networks are complex structures described by many different variables. Daganzo and Geroliminis (2008) uses variational theory to approximate the MFD of a street network. First they derive the cost function, which is an essential part of the variational theory, for a single street. Accordingly, by means of a similar method, also the cost function of a street network is determined.

2.6.2 A single street with no turns

In order to compute the relative capacity ("cost") function $r(u)$, this paper makes two assumptions:

- A street can have any number of time-invariant and/or time-dependent point bottlenecks with known capacities
- There exists a set of "valid" paths (observers average speed at any time interval is in the range $[-w, u_f]$)on the (t, x) plane starting from an abitrary point on the boundary at $t = 0$ and ending at a later time $t > 0$.

As a consequence, they derive that $R(u)$ is an upper bound to the average rate at which traffic can overtake any observer that travels with the average speed u. Furthermore, given that a ring's MFD is concave, they show that the flow is maximum the largest flow which can pass any observer moving with average speed u. In mathematics:

$$q = \inf_u \{ku + R(u)\} \tag{2.31}$$

So if one tests all "valid" paths, one can establish the MFD of a ring street, see figure 2.8.

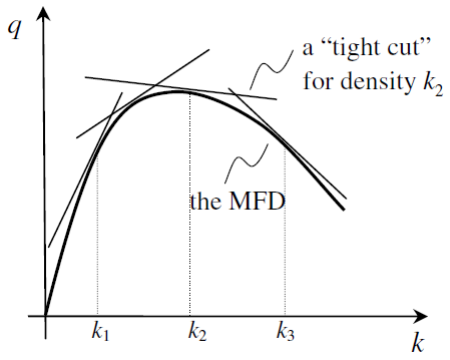

Figure 2.8: The MFD defined by a 1-parameter family of "cuts" Daganzo and Geroliminis (2008)

Yet, since testing all "valid" paths is tedious, Daganzo and Geroliminis propose the use of three families of "practical cuts" which jointly bound the MFD from above. These cuts are:

1. Family 1: The stationary observer standing at the most constrained intersection. This cut provides:

$$q \leq q_B = \min_i \{s_i G_i / C_i\} \tag{2.32}$$

2. Family 2: Observers which move forward at speed $u_{freeflow}$, except when detailed by a red phase.

$$q \leq ku(\epsilon) + \sum_i s_i f_i(\epsilon) \tag{2.33}$$

   where s_i is the passing rate of the observer during the fraction of time f_i that it spends artificially stoped in green phases.

3. Family 3: Observers traveling opposite to the direction of the flow with speed w, and also stop at red phases.

$$q \leq -kw(\epsilon) + \sum_i [s_i b_i(\epsilon) + r_i h_i(\epsilon)] \tag{2.34}$$

   where b_i is the fraction of time the observer spends at in the extended red phases of intersection i, and h_i si the fraction of time it spends moving towards i.

For homogeneous street (networks), these cuts can be further specified using less parameters. See the paper for the exact formulation in this case.

2.6.3 networks

Two new issues arise if one would like to adopt this approach to compute the MFD of a network, being:

1. Vehicles actually do not all behave in a similar manner. The paper shows that also in this case a bound can be found for the average network flow. This network level bound $T(K)$ is:

$$T(k) = \max\{\sum_i T_i(k_i)\, l_i\ D | \sum_i (k_i)\, l_i\ D = k; 0 \leq k_i \leq \kappa_i\} \geq q \tag{2.35}$$

   if the network has (i) slow-varying and distributed demand, (ii) a redundant network, (iii) a homogeneous netwrok with similar links and (iv) links with an approximate FD that is not significantly affected by turning movements, They show that the flow is approximate $T_i(k_i)$.

2. Stochastic effects in the network induced by turns and trip ends. Conclusions about this issue are not clear..

Daganzo and Geroliminis show by means of a case study that their approach provides a reasonably correct estimation of the MFD.

Chapter 3

An Introduction to Node Models

Lin Xiao and Jeroen P.T. van der Gun

3.1 Introduction

Links and nodes are the essential elements of network topology. No matter we consider people or goods transported in physical network or data delivered in logical network, the structure of network is determined by the arrangement of these components. Each network, regardless of its complexity, can be schematized as an abstract topological network with links and nodes only. So does the traffic network.

In traffic simulation, road network are generally built up by connected links and nodes. Links are considered as the roads where the vehicles travel on. And nodes are the intersections of these links, which passing flows from and to different segments. Therefore, the connected links and nodes form a network to support vehicles' movements from each origin to various destinations.

Important for traffic simulation is the ability to mimic the propagation of vehicle flows along the network. That is the dynamic network loading (DNL) model, which, together with a route choice model, forms a dynamic traffic assignment (DTA) model. The node model is part of DNL model and plays significant role in simulating the flow's motion.

The node model is the mechanism of receiving flow and sending flow at a node. It reflects the movement of flow from upstream to downstream links and determines how much flow can be actually carried out. Flow is constrained by traffic demand and supply. In general, traffic demand are the flows which intend to move from incoming links to outgoing links. And traffic supply is defined as the maximum flows that outgoing links can receive. Because of the imbalance between demand and supply, actual flows are not as same as the demand flows. Therefore, node model is expected to be a distribution process of traffic flow, transforming an OD matrix into a flow matrix (see fig 3.1).

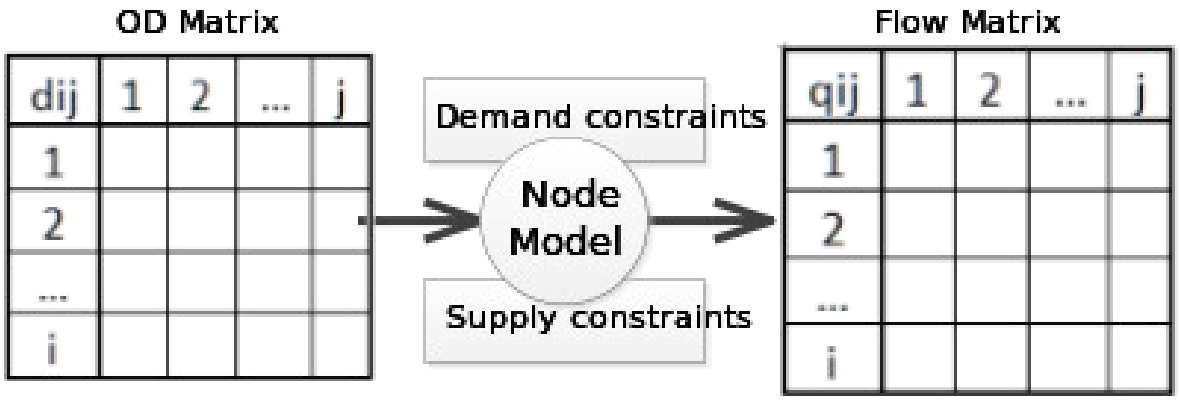

Figure 3.1: OD matrix and flow matrix as input and output of the node model

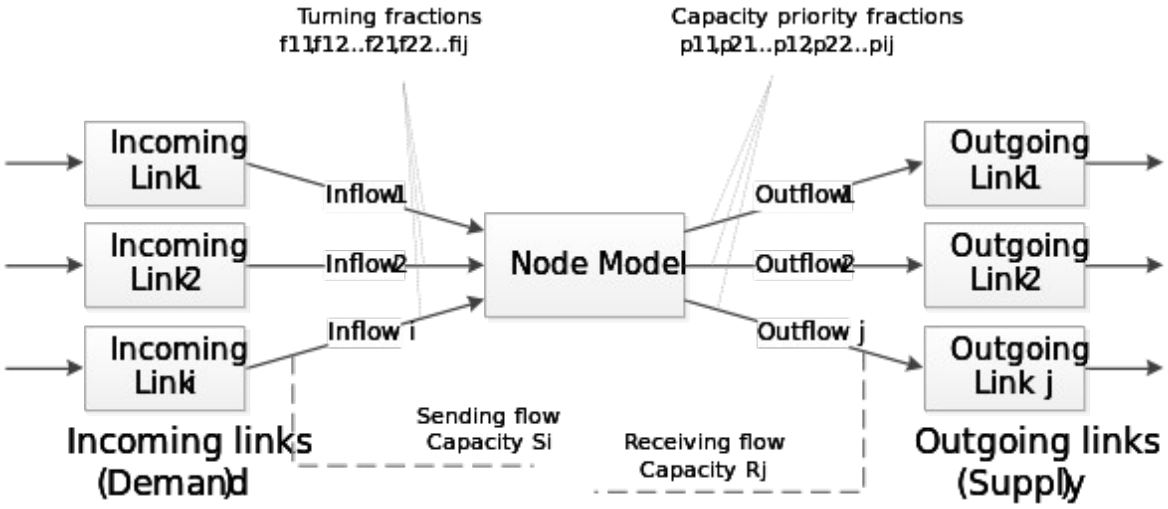

Figure 3.2: Node model structure

The general structure of a node model is shown in fig 3.2. Each node model is connected with one or several incoming links (1,2... i) and outgoing links (1,2... j), which sending and receiving flow from and to node model. Incoming link i holds the flows that desire to move to each outgoing link, as the d_{ij} in row i of OD matrix. Sum of these demands are the total demand of incoming link i. But due to node constraints, actual flows q_{ij} may not as same as the demand, represented by the sending flow S_i. We thus have a demand constraint:

$$\text{Inflow}_i = \sum_j q_{ij} \leq S_j \tag{3.1}$$

Outgoing link j receives flows from each incoming link, as the q_{ij} in j column of flow matrix. Outflow j is the total flows towards outgoing link j, which is restricted by the ability of the downstream link to accept more vehicles, indicated by receiving flow R_j. We thus have a supply constraint:

$$\text{Outflow}_j = \sum_i q_{ij} \leq R_j \tag{3.2}$$

The turning fraction f_{ij} is defined as the ratio of flow q_{ij} over inflow i, which indicate the destination distribution of node inflow. From the opposite side, assuming the capacity of outgoing links is fully used, we can define priority fraction p_{ij} as the ratio of demand d_{ij} over capacity R_j. It is used for reflecting the distribution of capacity for each incoming flow. They can be written as:

$$f_{ij} = \frac{q_{ij}}{\sum_j qij} \tag{3.3}$$

$$p_{ij} = \frac{d_{ij}}{R_j} \tag{3.4}$$

In Tampère et al. (2011) two functions of node model are described. First, node model imposes constraints on outflow of incoming links because imbalanced demand and supply. If node demand is larger than node supply, not all the demand flow can be satisfied and flow is then limited to the minimum flow that can be achieved. Therefore outflow of upstream links are constrained.

The second function of node model is to see consistency between demand and supply constraints. From equations (1) and (2), we know that each flow q_{ij} are constrained by both two constraints. But the constraints imposed only on link inflow and outflow leads to the interactions between each flows. It means any change of q_{ij} may lead to changes of the other flow. Thus the node model distributes flow under the guarantee of consistency, and it becomes a flow optimization problem with demand and supply constraints.

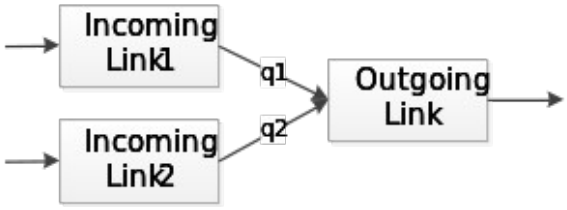

Figure 3.3: Merge model

Two key papers in node model are introduced here to get further details about how the node model works. The paper on the Cell Transmission Model by Daganzo (1995) is introduced first with its merge and diverge models. Then the work of Tampère et al. (2011) is presented with the concept of capacity-proportional priorities.

3.2 Node models in the Cell Transmission Model

Daganzo (1995) first defines the two basic forms of a node model he considers: a merge model and a diverge model. Restricting the attention to networks with three-legged junctions, Daganzo described merge model as two links enter the cell and only one link leave it, while diverge model has one incoming link and two outgoing links. Valid node representations are outlined in paper and stressed that links can only belong to one of merge model and diverge model. Junctions with more legs are not allowed here.

3.2.1 Merge model

Considering three-legged junction, we have two incoming links (i=1,2) and one outgoing link (j=1) in merge model (fig 3.3). Then the objective of the merge model is to maximize the total flow q into the outgoing link:

$$\max q = q_i + q_j \tag{3.5}$$

Demand constraints of the incoming links are S_1 and S_2 respectively. The outgoing link has the supply constraint as R, with capacity priority fractions p_1 and p_2 for link 1 and link 2($p_1 + p_2 = 1$). Then, because of demand and supply constraints, the flow should satisfy:

$$q_1 \leq S_1 \tag{3.6}$$
$$q_2 \leq S_2 \tag{3.7}$$
$$q_1 + q_2 \leq R \tag{3.8}$$

It is straightforward that if $q_1 + q_2 \leq R$, then $q = q_1 + q_2$.Traffic state is free flow on both incoming links. This case is marked as (i). If $q_1 + q_2 > R$, vehicles accumulate in at least one of the incoming links. Then two sub-conditions are considered. (ii) One link is congested and the other link is free flow, then solutions are $q_1 = S_1$, $q_2 = R - S_1$, $q = R$. (iii) Both links are congested, then flows are distributed by priority fractions as $q_1 = p_1 R$, $q_2 = p_2 R$, $q = R$. . With the help of fig 3.4, these three cases are illustrated. The rectangle in figure represents the set of flows that satisfy each demand constraint. Points below downslope line $q_1 + q_2 = R$ meet the requirement of supply constraint. Line $\frac{q1}{q_2} = \frac{p_1}{p_2}$ represents the ratio of flows if the supply is not exhausted.

In fig 3.4, case (i), the descending line does not cross the rectangular area, which means all flows satisfying demand constraints also satisfy the supply constraint. Therefore the flows are determined by minimum restrictions – the demand constraints – only, and the maximum flow should be the intersection of two demand boundaries (point P). Here free flow is expected in both upstream links. In case (iii) the descending line intersects the ascending line within the rectangle. Because the outgoing flow is distributed according to the capacity priorities which are proportional to the demands, to achieve maximum total flow, the solution is the intersection of these two lines, (point Q). Finally, in case (ii), the descending line

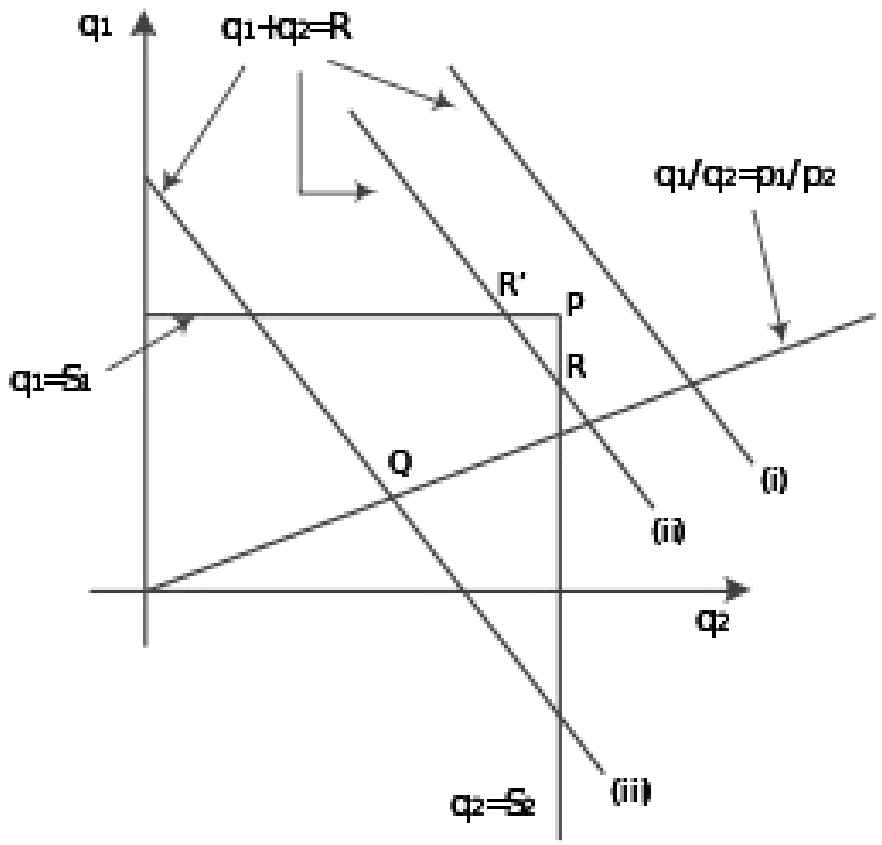

Figure 3.4: Diagram of feasible flows for a merge junction

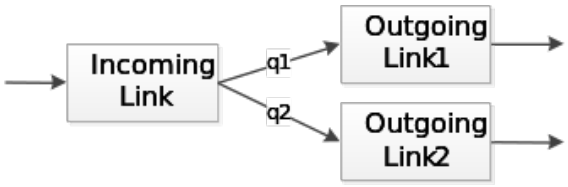

Figure 3.5: Diverge model

is not totally above rectangle but it intersects the ascending line outside the rectangle. The intersection is not within the feasible flow set, indicating that no flows with these priorities can reach the maximum total flow R. Within the rectangle representing demand constraints, the maximum can be only achieved along line segment RR'. And since the supply on one of the incoming link is exhausted in case (ii), a solution can only achieved on one of the demand boundaries, leaving us with only the two endpoints of this line segment. The point is the point closest to the ascending line (point R), as q_2 should not be further reduced than required by its demand constraint.

Therefore, in case (ii) and (iii), it can be concluded that the solution point for this optimization problem is the middle point of the three intersection points of the descending supply constraint line. This solution is formulated as:

$$q_1 = \text{mid}\{[S_1, R - S_2, p_1 R\} \tag{3.9}$$
$$q_1 = \text{mid}\{S_2, R - S_2, p_2 R\} \tag{3.10}$$

if $q_1 + q_2 > R$.

3.2.2 Diverge model

In the diverge model, there is one incoming link (i=1) and two outgoing links (j=1,2). The objective is to maximize the total outflow q of the incoming link, and hence to maximize the total inflow of the downstream links. Demand constraint for incoming link is S, supply constraints are R_1 and R_2 respectively. With the turning fractions given, flows should satisfy

$$q_1 = f_1 q \leq R_1 \tag{3.11}$$
$$q_2 = f_2 q \leq R_2 \tag{3.12}$$
$$q_1 + q_2 \leq S \tag{3.13}$$

It is assumed that a vehicle unable to exit the incoming link hinder the vehicle behind, which implies that the first-in-first-out rule (FIFO) is applied in the diverge model, which is also known as conservation of turning fractions (CTF) in this context. Consequently, if the flow going to link 1 is constrained, the flow into link 2 is constrained accordingly. The solution of diverge model can be formulated as:

$$q = \min\{S, R_1/f_1, R_2/f_2\} \tag{3.14}$$

Note that the turning fractions can be specified explicitly, potentially as a function of time, but can also be derived by modelling traffic flows heterogeneously. In Daganzo (1995), traffic is for example disaggregated by destination, and we can then specify the turning fractions per destination. E.g. Yperman (2007) further generalizes this into the concept of multi-commodity modelling. Van der Gun et al. (2015) for example instead use disaggregation by origin to specify turning fractions.

3.3 General node models

The previously discussed node models only apply to simple merges and diverges. General node models with an arbitrary number of incoming and outgoing links are a lot more difficult to formulate. Tampère et al. (2011) cite Bliemer (2007) as a typical general model for which they provide a numerical example.

An important problem with the Bliemer model, that also occurs in the previously described Daganzo (1995) merge model, is that it does not satisfy the so-called invariance principle. Introduced by Lebacque (2005), this principle states that flows across an intersection should not change during an infinitesimal time step if demand and supply remain constant. This principle turns out to be violated for any node model based on priority fractions as defined in Equation (3): here, congestion may be formed on an incoming link, resulting in an increased demand after an infinitesimal time step and consequently also a higher priority and higher flow. This creates unstable, alternating flows in the dynamic network loading process, which is not desirable.

Before formulating a node model themselves, Tampére et al. first formulate the list of requirements it should satisfy:

- Support arbitrary numbers of incoming and outgoing links.
- Maximize flows. Similar to the difference between user-equilibrium and system-optimal assignment, this may well be a local maximum because drivers approaching an intersection are selfish. This selfish behavior is modelled by supply constraint interaction rules (SCIRs).
- Produce non-negative flows.
- Conserve vehicles.
- Satisfy demand and supply constraints of the connected links. There may be additional internal node supply constraints, e.g. to account for crossing flows that do not have an incoming or outgoing link in common. The previously mentioned SCIRs should also be satisfied.
- Conserve turning fractions (CTF).
- Satisfy the invariance principle.

Then, they mathematically formulate node models for unsignalized and signalized intersections satisfying these requirements. The employed solution to satisfy the invariance principle is to define priorities proportional to the capacities (i.e. maximum demands) of incoming links rather than the current demands of incoming links. The unsignalized intersection model does not include node supply constraints, whereas the signalized intersection model includes node supply constraints representing the green phases of the traffic lights. The solution algorithm is too complex to describe here, so the interested reader is referred to the paper for details. Despite its complexity, it is guaranteed to converge to the solution within a fixed, finite number of iterations.

3.4 Discussion

One may wonder whether there exist other approaches to solve a general node model. As an alternative to the Tampère et al. (2011) algorithm, Flötteröd and Rohde (2011) propose an incremental procedure and extend that to include node supply constraints, finding the same solution in a different way. There seems plenty of work to do to operationalize and validate node models for all kinds of complex intersection shapes that exist in reality, including e.g. roundabouts. For example, Smits et al. (2015) show that while the invariance principle can be satisfied without assuming capacity-proportional priorities, the solution algorithms of the such node models may not fully converge within finite time.

Although historically, as in Daganzo (1995), node modelling was only "part 2" after link modelling based on fundamentals like Kinematic Wave Theory, it now appears that node modelling has become more of a fundamental research area itself.

We also like to note there is an analogy between the above discussion of node models and the lane change model of Laval and Daganzo (2006b). That lane change model can be expressed in terms of a network, where each link represents a lane for the length of some segment and each node between segments represents an opportunity to change lanes. With this in mind, their lane change model is actually a node model with demand-proportional priorities.

Chapter 4

The network fundamental diagram

Goof S. van de Weg en Victor L. Knoop

Preface

This document provides an introduction to the Network Fundamental Diagram (NFD). By reading this document one should gain insight into:

- the main idea of the NFD
- the purpose of the NFD;
- the existence of the NFD;
- approaches to model the NFD;
- approaches to model traffic dynamics using the NFD;
- traffic control using the NFD;
- determining sub-networks that can be modelled using the NFD.

4.1 The concept of the network fundamental diagram

To describe the traffic flow in the city center of Delft as shown in Figure 4.1, there are various options. For instance, one could describe the traffic using a microscopic model – i.e. by modeling individual vehicle dynamics – or using a macroscopic model – i.e. by modeling traffic flows at a link level. However, both these approaches will cost a lot of computation time, especially, when the interaction between a lot of regions or neighborhoods of the entire city of Delft have to be modeled. In order to address this problem, researchers have been searching for a way to describe traffic at an aggregated level – i.e. by considering the dynamics of the entire region using a simple function – for almost half a century according to Geroliminis and Daganzo (2008).

About a decade ago the concept of the NFD was suggested by Daganzo (2005d, 2007b). The NFD describes the out-flow out of an urban network as a function of the accumulation – i.e. the number of vehicles in an urban road network – and the in-flow to the network. Hence, it is a function that describes the dynamics of an entire urban area using a simple relation. An example of the NFD is illustrated in Figure 4.2. When the number of vehicles stays to the left of the critical accumulation, the network production increases with the accumulation. However, when the accumulation exceeds the critical accumulation the production starts decreasing again. Thus, using this simple relation, already some important dynamics of the operation of a traffic network can be modeled.

Daganzo (2005d, 2007b) argued that this relation could describe the accumulation over time when in-flow to the network slowly changed over time and the congestion in the network would be homogeneously distributed. Using microscopic simulation the existence of the function was shown and also perimeter

Figure 4.1: Google maps screens-shot of the Delft city center.

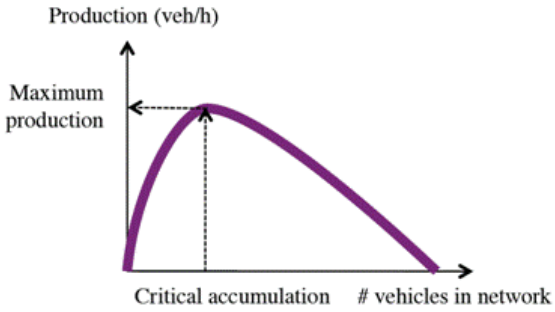

Figure 4.2: An example of the fundamental relation of network production.

control – i.e. limiting the flow into the region such that the accumulation stays to the left of the critical accumulation – was tested.

So, getting back to the example, using the NFD it is possible to describe the aggregate dynamics of the entire Delft city center. Compared with other traffic flow models, such as macroscopic or microscopic models, this description is much more simple. Hence, the traffic flow model will become much faster which is beneficial when very large networks, e.g. the entire city of Delft, have to be modeled.

Of course, the fact that this model is computationally efficient comes at a cost. This cost is represented by the fact that a lot of detail of the traffic dynamics is lost. In order to get an idea of this detail lets look at the various factors that influence the traffic dynamics, one can think of: the road lay-out, the driving behavior, the traffic demand and routing, the weather, incidents, traffic control strategies, etcetera. Apart from that, phenomena in the traffic dynamics are not necessarily homogeously distributed. For instance, Knoop et al. (2013) show that congestion is usually hetereogeneously distributed.

This document further explores the literature on the NFD. First it will present look at some experimental findings on the existence of the NFD, followed by an approach to analytically derive the NFD. Next, section 4.4 describes a model based on the NFD. After that section 4.5 discusses traffic control using the NFD. The final section presents the approach to partition an area into several sub-regions such that the NFD can be used to describe the traffic of the total area.

4.2 The existence of the network fundamental diagram

Geroliminis and Daganzo (2008) studied the existence of the NFD for downtown Yokohama, Japan, using real data. GPS data was gathered from 140 taxis, and vehicle counts and occupancy measurements from 500 ultrasonic and inductive loop detectors were collected. Comparing the fundamental diagrams of two different loop detectors showed a lot of scatter within the individual fundamental diagrams and two clearly distinguishable NFDs. However, when aggregating all the data – i.e. by taking the average over the flow and occupancy – and plotting the corresponding density and flow, a clear NFD with little scatter appeared as shown in Figure 4.3 was found. The approach to aggregate the data was as follows:

$$q = \sum_i \frac{q_i l_i}{\sum_i l_i}\,, \tag{4.1}$$

$$o = \sum_i \frac{o_i l_i}{\sum_i l_i}\,, \tag{4.2}$$

$$\tag{4.3}$$

where q_i is the flow in link i, o_i the occupancy measured in link i, and l_i the length of link i.

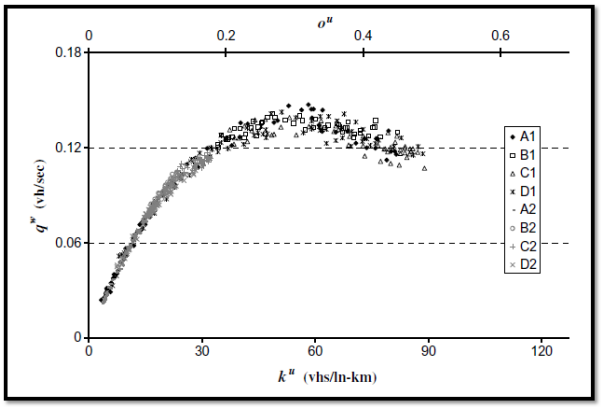

Figure 4.3: The NFD of downtown yokohama of Geroliminis and Daganzo (2008).

Apart from that, the authors used the taxi data in order to investigate the relationship between production and trip completion rate. According to the theory this relationship should be linear. Hence, the authors use the taxi data in order to estimate the total trip completion rate of the entire downtown area. The authors find a linear relation-ship with low scatter.

4.3 An analytic derivation of the network fundamental diagram

The previous section indicated that it is possible to estimate a NFD from real traffic data. However, it would also be useful if the NFD could be derived analytically. The reason why this is helpful is that it could be (potentially) understood what affects aggregate network performance and how a control strategy affects the aggregate network performance.

Daganzo and Geroliminis (2008) proposed an approach to derive a NFD based on the road geometry and signal timing. The authors note though that this work should be seen as a first step to derive such a method. Ideally, a 'universal recipe that can be used for all signal-controlled networks' should be developed. The approach is based on variational theory and here we will try to detail the essence of the approach. The interested reader is referred to Daganzo and Geroliminis (2008) for the full explanation of the approach. The approach consists of several steps, first it is detailed how the capacity of only one road can be determined using variational theory, next it is explained how this can be used to estimate a NFD, finally the approach is applied to data from a micro-simulation and real data from downtown Yokohama, Japan.

4.3.1 Determing a street's capacity using variational theory

Variational theory is the theory that deals with the problem of finding the path between two points with the lowest costs. Why this can be used to estimate the capacity of a street will be explained in the remainder of this section. Note that a street's capacity is defined as the maximum number of vehicles that can pass a certain point in time. Let us denote this capacity with $R(0)$ which indicates the rate at which vehicles pass an observer which travels with an *average* speed of 0 km/h. An example of such an observer is indicated with the red dashed line from point A at time t_A to point B at time t_B in Figure 4.4. This figure illustrates a time-space plot of a road with three intersections. The driving direction is from bottom to top and the horizontal solid lines indicate that a traffic light is showing a red light

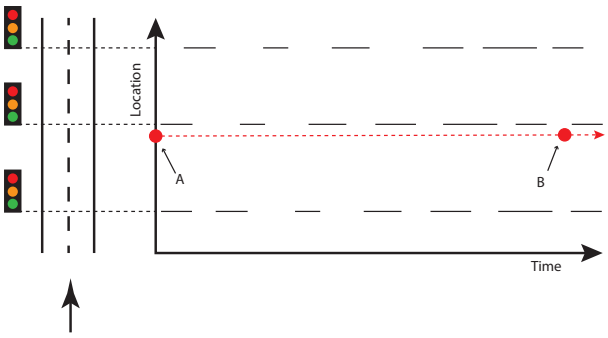

Figure 4.4: A time-space representation of a street with three controlled intersection. The dashed line from A to B indicate an observer which travels with an average speed of 0 km/h. The solid horizontal lines indicate that the traffic light is red.

To illustrate this, consider the time-space plot in Figure 4.4. As can be seen, it consists of a street with several controlled intersections of which the signal timing is known, the horizontal lines indicate that a red signal blocks the stream. When it is assumed that flow at the downstream end of the street matches the flow at the upstream end, variational theory can be applied to find the capacity of the street. The street's capacity is denoted with $R(0)$ which is equivalent to the number of vehicles that passes an observer that travels with an *average* speed of 0 km/h. An example of such an observer is indicated with the red dashed line from point A at time t_A to point B at time t_B in Figure 4.5.

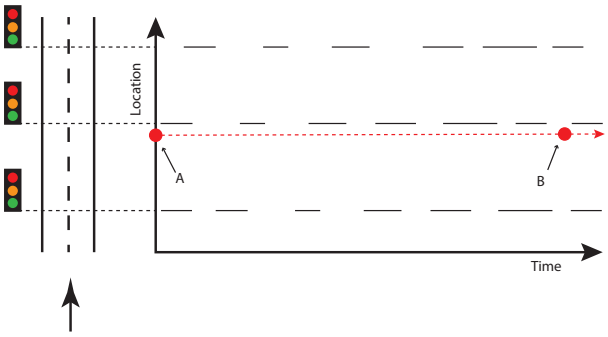

Figure 4.5: A time-space representation of a street with three controlled intersection. The dashed line from A to B indicate an observer which travels with an average speed of 0 km/h.

The question now is: how do we determine the number of vehicles passing the observer from the time from A to B? This is where variational theory comes in. Please note that we assume here that the flow at the downstream end of the street matches the flow at the upstream end. Now, the main concept is as follows. When we draw a trajectory of an observer through the time-space plot, the number of vehicles $r(u)$ that passes the observer at a given location is given by the following equation:

$$r(u) = \sup\{Q(k) - ku | \forall k \in [0, \kappa]\}\,, \tag{4.4}$$

where $Q(k)$ is the fundamental diagram – i.e. the relation between density and flow –, k is the average density on the road and u is the speed of the observer. κ indicates the maximum density, i.e. the jam

density. An example of such a trajectory with average speed 0 is indicated with the curved line from point A to B in Figure 4.6. Since we know the speed u of the observer at any time instance, the maximum number of vehicles passing the observer at every time instance can be computed using Eq. (4.4). Note that an infinite number of paths $\mathcal{P}$ from A to B can be found and for all these paths the associated costs $\Delta(\mathcal{P})$ can be computed. For all these paths it should hold that the speed is between the backward wave speed $-w$, and the free-flow speed u_f.

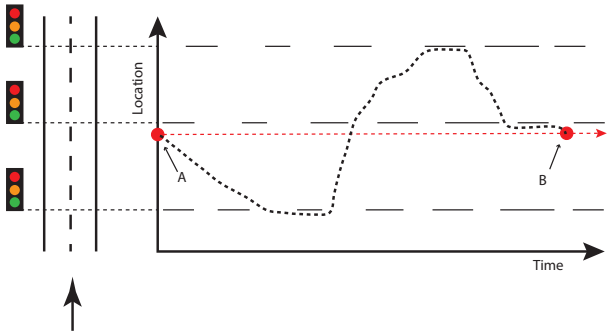

Figure 4.6: The dashed curved line indicates an example of a path of an observer that has an average speed of 0 km/h. Note that it is possible to define an infinite number of paths from A to B.

Now, the passing rate of any observer with average speed $R(u)$ is given by the path with the lowest costs:

$$R(u) = \frac{\lim_{t_0 \to \infty} \inf_{\mathcal{P}} \{\Delta(\mathcal{P}) : u_{\mathcal{P}} = U)\}}{t_0}. \tag{4.5}$$

Following this approach the street's capacity $R(0)$ can be determined. How to find this minimum in Eq. (4.5) is not detailed in the paper, it is an optimization/dynamic programming problem which should be solved using a shortest path algorithm.

In the paper of Daganzo and Geroliminis (2008) it is assumed that a triangular fundamental diagram can be used to describe the dynamics of the traffic. This fundamental diagram is characterized by the optimal density k_0, the free flow speed u_f, the jam density κ, the backward wave speed w, the capacity q_m, and the maximum passing rate r. Also, when a red-light is shown, the flow $q_B = 0$ at the bottleneck location is zero and when a green-light is shown the flow $q_B = s$ is equal to the saturation flow. Under these assumptions **?** show that 'an optimal path always exists that is piece-wise linear: either following an intersection line or else slanting up or down with slow u_f or $-w$.' The possible directions of the path and the corresponding passing rates are shown in Figure 4.7 from Daganzo and Geroliminis (2008).

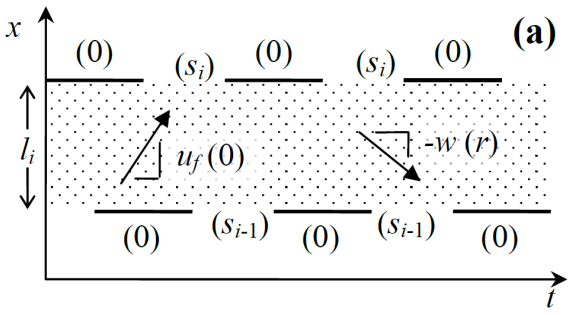

Figure 4.7: Illustration from Daganzo and Geroliminis (2008) showing the different speeds with which an observer can travel, and the corresponding passing rates of traffic.

4.3.2 Determining the network fundamental diagram using variational theory

The previous section detailed the approach to determine the capacity of a road using variational theory. In order to understand this section the minimum concept that should be understood is that by means of variational theory it is possible to determine the passing rate $R(u)$ with which traffic passes an observer traveling with the average speed u. Now, a next step is made to determine a NFD using variational theory.

Note that the capacity of a street is defined as the number of vehicles n_0 that can pass a moving observer which travels with an average speed of 0 km/h. Thus, we are interested in the flow $q(k) = n_0/t_0$ passing an observer that travels with a speed of 0 km/h over a time t_0. Given an observer that travels with an average speed of u the number of vehicles n_0 is given by:

$$n_0 = kut_0 \pm kL + R(u)t_0\,, \tag{4.6}$$

where L is the length of the link. Here we will give an intuitive explanation of this equation, to this end consider Figure 4.8. In this figure, the blue dashed line with slope u indicates the observer. In this case, we are interested in n_0 i.e. the number of vehicles that passed the stationary – i.e. with speed 0 km/h – observer assuming that the vehicle number at the begin point A equals zero. During the time t_0 the observer travels a distance ut_0, hence the point C lies a distance of ut_0 downstream of A. Considering an average density on the street of k, the number of vehicles between A and C is then kut_0. Next, the number of vehicles that passes the observer during the time t_0 is equal to $R(u)t_0$. So, the number of vehicles at n_0 is $kut_0 + R(u)t_0$. However, this holds when we assume that the number of vehicles in A is 0. This number can have a maximum absolute error of kL hence, the $\pm kL$ term in Eq. (4.6).

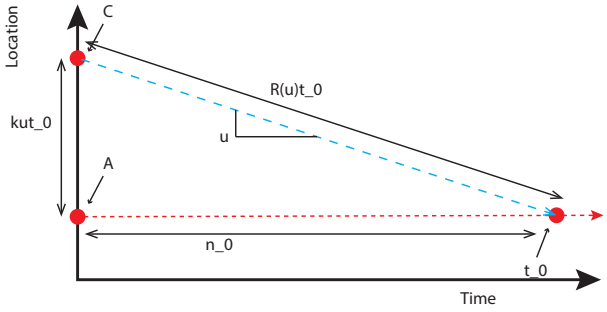

Figure 4.8: Determining the number n_0 using an observer traveling with average speed u.

When dividing both sides of Eq. (4.6) by t_0 and letting $t_0 \to \infty$ the constant term kL drops out and we get the following expression:

$$q(k) = ku + R(u)\,. \tag{4.7}$$

When plotting this expression in a density flow diagram we get a line starting at $R(u)$ with a slope of u as illustrated in Figure 4.9. When determining this line for every observer, and taking the minimum of these lines we end up with a fundamental diagram:

$$q = \inf_u \{ku + R(u)\}\,. \tag{4.8}$$

This is illustrated in Figure 4.10. In this figure, some of the observers are indicated with dashed lines. The solid curve indicated the fundamental diagram that follows from Eq. (4.8). One has to imagine that the more of these observers are computed, the smoother the fundamental diagram becomes.

The above approach implies that for every value of k, all the values of u have to be assessed. However, this is computationally very demanding. Therefore, the method is simplified. Instead of analyzing the density for every possible average observer speed, only certain families of observer speeds are evaluated. These families originate from the free-flow speed, the backward wave speed and the zero speed observer. By assessing the capacity for these families of speeds, equations that relate the density and the flow for a certain speed are known. For a detailed description of (the derivation of these) equations, the interested

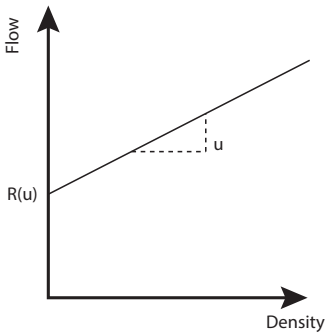

Figure 4.9: Representation of the capacity flow related to an observer traveling with average speed u.

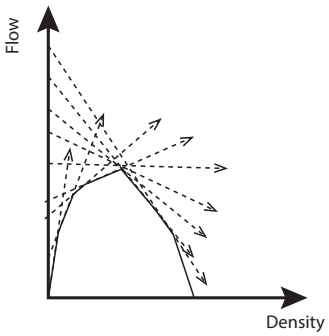

Figure 4.10: The line

reader is referred to Daganzo and Geroliminis (2008). Using this approach, the NFD is approximated by means of a number of affine lines which provide an upper bound to the NFD. This is shown in Figure 4.11.

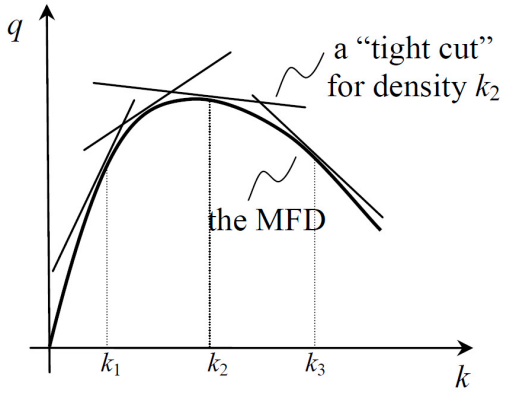

Figure 4.11: Illustration from Daganzo and Geroliminis (2008) showing the affine lines that provide an upper bound to the NFD.

4.3.3 Application to real traffic networks

The final step is the application of the theory to real traffic networks to get the network fundamental diagram $T(k)$. This is realized by adding up the weighted fundamental diagrams of every link in the network. The fundamental diagrams are weighted according to the fraction of the link length l_i divided by the total network length D. The task is to combine the fundamental diagrams $T_i(k_i)$ of all the

individual links given the accumulation k in the entire network. Daganzo and Geroliminis (2008) note that the link fundamental diagrams provide an upper-bound to the actual link fundamental diagrams. Hence, combining the fundamental diagrams should result in an upper-bound of the network fundamental diagram. To accommodate this, the link densities k_i are adjusted in such a way that the sum of the link densities is equal to the network accumulation and the corresponding sum of the link productions $T_i(k_i)$ is maximized:

$$T(k) = \max \left\{ \sum_i T_i(k_i) l_i / D \middle| \sum_i k_i l_i / D = k;\ 0 \leq k_i \leq \kappa_i \right\} \tag{4.9}$$

The authors argue that when all the links are similar the optimum is found for $k_i \cong k$. Also, the authors note that under a number of conditions this NFD not only provides an upper bound, but a good approximation of the actual NFD. These conditions are: 1) the demand should be slow-varying and distributed, 2) the network should be redundant – e.g. an irregular grid made up of block – such that drivers have many route choices and that most links are on many desirable routes, 3) the network should be homogeneous with similar links, and 4) the fundamental diagrams of the individual links should not be significantly affected by turning movements.

Now, in order to apply Eq. (4.9) to a real network several simplifications have to be made. For instance, when applying the approach to the Yokohama network it is assumed that the entire network can be constructed of identical links. Also, assumptions are made on the signal timing. For more detail the reader is referred to Daganzo and Geroliminis (2008). The result of the approximation compared with the actual NFD is shown in Figure 4.12

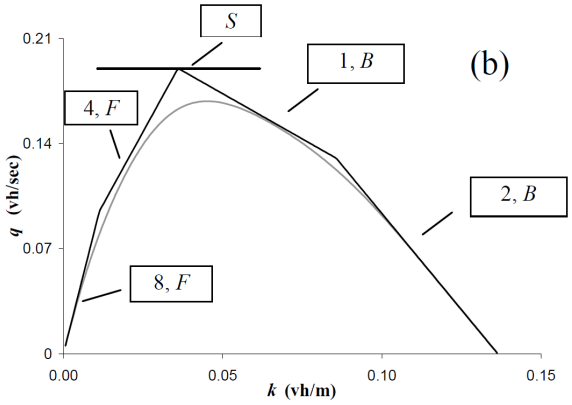

Figure 4.12: The NFD of the Yokohama network compared to the piece-wise affine analytically derived NFD of Daganzo and Geroliminis (2008).

4.4 Traffic flow modeling using the network fundamental diagram

The previous sections detailed how a NFD can be estimated from data, and how a NFD can be derived from network geometry and signal timing. The next step is to model the dynamics of traffic using the NFD. Knoop and Hoogendoorn (2014) proposed the network transmission model in which the dynamics of the traffic of a large urban area is described by the flows between sub-areas of which a NFD is known. Here we assume that the reader is familiar with some basic traffic flow modeling concepts.

Knoop and Hoogendoorn (2014) divide a network into several sub-networks. A NFD $P(K)$ relating production P to accumulation K is known for every sub-network. Next, the network transmission model describes, based on the demands and routes of traffic how the accumulation K of every sub-network evolves over time.

The approach to determine the flows between sub-networks is similar to the cell transmission model. The key is that the demand D and supply S of two neighboring sub-networks have to be determined.

The supply is at the capacity if the accumulation in the receiving cell is below the critical accumulation and equal to the NFD if it exceeds the critical accumulation:

$$S = \begin{cases} P_{crit} & \text{if } K \leq K_{crit} \\ P(K) & \text{if } K > K_{crit}\,. \end{cases} \tag{4.10}$$

The demand, i.e. the flow out of a sub-network is equal to the production in the cell:

$$D = P(K)\,. \tag{4.11}$$

This implies that when the accumulation exceeds the critical accumulation, the out-flow decreases.

Note that the demand to a neighboring sub-network depends on the number of vehicles that want to go to the neighboring region. The model accounts for the fraction of traffic going towards a certain region, for all the details the reader is referred to Knoop and Hoogendoorn (2014). Note that in the end, the flow that can travel from one region to another is the minimum of the demand and supply.

4.5 Traffic control using the network fundamental diagram

One of the applications of the NFD is perimeter control. With this term it is meant that the flows between various sub-networks or the in-flow to a single sub-network are controlled. Recently, various approaches for the control of traffic flows between sub-networks have been proposed that describe the dynamics of the sub-networks using the NFD. Typically, these control strategies aim at keeping the accumulation in the network below the critical accumulation. In so doing, the negative effects of the capacity drop are prevented. In the literature, two different control paradigms can be distinguished, namely (1) feedback control, Keyvan-Ekbatani et al. (2012); Aboudolas and Geroliminis (2013), and (2) MPC control, Geroliminis et al. (2013); Haddad et al. (2013); Hajiahmadi et al. (2013).

Feedback perimeter control

Keyvan-Ekbatani et al. (2012) propose a Proportional-Integral (PI) feedback control strategy. The aim of the control strategy is to adjust the in-flow $q_{in}(k)$ to the sub-network at time step k such that the total time spent $TTS(k)$ of the network reaches a preferred set-point $T\hat{T}S$. The following control law is proposed:

$$q_{in}(k) = q_{in}(k-1) - K_p(TTS(k) - TTS(k-1)) + K_I(T\hat{T}S - TTS(k))\,, \tag{4.12}$$

where K_p and K_I are the proportional and integral gain. The control law is evaluated using a microscopic simulation of the urban network of Chania, Greece. It is found that the average speed in the network increases with up to 40%.

Aboudolas and Geroliminis (2013) propose a linear multivariable feedback regulator that exploits the Macroscopic Fundamental Diagram (MFD) to balance the traffic demand between regions. An important difference with Keyvan-Ekbatani et al. (2012) is that the control law takes the performances of different areas into account. The control objective is twofold, the accumulation of vehicles has to be distributed homogeneously over the different regions, and the rate of vehicles that is allowed to enter each reservoir has to be maintained at a desired set-point.

The authors describe the dynamics of the whole system as follows:

$$\Delta\dot{n}(t) = F\Delta n(t) + G\Delta\beta(t) + H\Delta(t)\,, \tag{4.13}$$

where n is a vector with the accumulations in every region, Δn is a vector with the difference $n_i - \hat{n}_i$ where $\hat{n}_i$ is the optimal accumulation of region i. Furthermore, $\Delta\beta(t)$ is the control deviations vector, i.e. the difference between the control action from the set-point, and $\Delta(t)$ is the demand deviations vector. F, G, and H, are the state, control, and demand matrices. Now, the control law is given by $\beta(k) = \hat{\beta} - K(n(k) - \hat{n})$, where K is found by solving the algebraic ricatti equation.

The control strategy is evaluated by means of a microscopic simulation of a three region network. Compared to the no-control situation, the throughput of the network improves by storing vehicles outside of the network.

MPC perimeter control

The MFD describes the relation between the production and accumulation of vehicles in a network. This relationship has first been described by Godfrey (1969). The MFD can be used for perimeter control by means of Model Predictive Control (MPC). Daganzo (2007b) proposed an optimal control policy that limits the inflow into a one-region area. Haddad et al. (2012) proposed a MPC for a two region area and Hajiahmadi et al. (2013) for a multi-region area. The approach of Hajiahmadi et al. (2013) uses the traffic model of Knoop and Hoogendoorn (2014) to describe the evolution of traffic in an entire network.

Haddad et al. (2012) developed a MPC based on the MFD in order to realize gating between two regions. The sampling time is equivalent to the cycle time of the intersection. The authors compare the MPC strategy with a greedy feedback control strategy using a simulation based on the fundamental diagram. The authors showed that the MPC strategy is superior to the greedy feedback control strategy. The authors note that splitting the network up in homogenous regions and applying gating can be a valuable control strategy.

Hajiahmadi et al. (2013) use the network transmission model in an optimization framework for perimeter control of a multi-region area. Route choices are explicitly modeled since traffic will travel across multiple regions. The control is realized by means of gating. The optimization minimizes the total delay in the entire network. The control strategy is evaluated by means of simulation of a 4 by 4 grid network of which the dynamics are simulated using the network transmission model. The algorithm is compared to the no control situation, a shortest path control algorithm and the MPC algorithm. The latter has the best performance. The authors note that 'the computation time is reasonable compared to other routing methods in the literature which are based on detailed modeling'.

Haddad et al. (2013) developed a MPC for perimeter control of a mixed urban and freeway network. The authors model the urban network by means of the MFD and the freeway network by means of the asymmetric cell transmission model. The control action is imposed by means of ramp metering and intersection controllers.

The authors compare various control strategies with the MPC control scheme. The authors use a traffic model to simulate the real world but do not report what kind of model they use. The MPC control scheme considers route choices where the user optimum is reached. In order to get the most efficient situation, the authors also control to the system optimum. In the most ideal case, consisting of centralized MPC that steers to the system optimum, the network performance increases with 10%. Essentially, the advantage of the MPC scheme is that the traffic is better distributed over the freeway and urban areas such that urban grid-lock is prevented, which results in a capacity drop. The authors do not provide insight into the computational complexity of the controller.

An important point of discussion of the above mentioned approaches is that it is assumed that the macroscopic fundamental diagram does not change when the control action is imposed. Moreover, an incident can change the macroscopic fundamental diagram which will affect the performance of this control strategy.

4.6 Sub-dividing a road network for use with the network fundamental diagram

An important question that should be asked when modeling/controlling a large urban road network using NFDs is where to put the boundaries between various sub-networks. During recent years, some algorithms have been developed to automatically partition road traffic networks. The approaches differ in the type of algorithm that is used, the partition approach, and the requirements for the partition that is needed. In this section several approaches will be detailed for the partitioning of road traffic networks which are developed for use with the NFD. In doing so, insight is gained into the requirements for partitioning a road traffic network, into the algorithms that can be used to partition networks, and into the limitations of these approaches.

4.6.1 Automatic partitioning of an urban network using the NFD

Ji and Geroliminis (2012) detail an approach to automatically partition an urban traffic network. The approach is inspired by image processing techniques. The partitions – also called clusters, and in this report sub-networks – that the authors create have to satisfy the following three (conflicting) properties:

(1) the variation in density in a sub-network should be as low as possible, (2) the number of sub-networks should be as low as possible, and (3) the sub-networks should be spatially compact. The reason for these requirements are that control strategies for the distribution of traffic among sub-networks based on the NFD can be developed, and that the complexity of the sub-networks controllers stays limited. To this end, the traffic in a sub-network should be as homogeneous as possible, and the number of sub-networks should be as low as possible in order to reduce the computational complexity of the control problem. On the other hand, the sub-networks should be spatially compact, such that effective control strategies within the sub-networks can be developed.

The paper details a three step approach to create the sub-networks. First, the Ncut algorithm is used to partition the network. Next, sub-networks are merged, and finally links at the boundaries of the sub-networks are moved from one sub-network to another until no more reduction in the variation of the density in the sub-networks can be achieved. The paper introduces metrics to assess whether the partitioning approach is optimal. The approach is applied to simulation data of the San Francisco network. It is shown that the method out-performs other partitioning approaches in terms of stability and reliability of the solution. The partitioning algorithm takes a few seconds to produce the sub-networks making it possible to dynamically change the sub-networks.

4.6.2 Partitioning to minimize the flow between sub-networks

The goal of the partitioning approach of Etemadnia et al. (2014) is to make sure that the flows between sub-networks are as small as possible. The authors argue that this reduces the need for traffic management schemes which require intensive communication and coordination among various controllers. Also, the sub-networks should be spatially connected meaning that intersections in a sub-network have to be adjacent to each other and the partitioning should result in a balance in the sense that the amount of traffic management activities which are required by each local controller is similar. This implies that sub-networks should be relatively similar in size, and/or they are similar in the amount of traffic served.

The objective is the minimization of flows between sub-networks. The objective function is constrained such that the goals mentioned in the previous paragraph can be achieved. Since the optimal control problem is NP-hard, two heuristic methodologies are developed to partition the network. Similar to Ji and Geroliminis (2012) the authors use cut algorithms. The interested reader is referred to Etemadnia et al. (2014) for more details on their approach. The approach is evaluated for various network set-ups, such as, a grid network, a circular network, and the Dallas, USA, network. In order to compute the partitions (static) link flow data is needed. The authors show that their approach is close to the optimal solution while requiring very limited computation time. It is also shown that different partitions might be needed for different flow patterns. Even for large networks, one of the approaches can determine the partitions in 4 seconds.

4.6.3 Automatic partitioning using Laplacian eigenvectors

Ma et al. (2009) developed a partition algorithm as well. The algorithm has to satisfy three requirements: intersections with a strong relationship should be grouped into the same sub-network and vice versa, and the number of intersections within every sub-network should, ideally, be equal such that the computational burden of the sub-network controllers is similar. The partition problem is formulated as an optimization problem which is NP-hard. The sum of the weight of the edges between different sub-networks is minimized, and a lower and upper bound to the number of intersections in a partition is given. The weight is given by the correlation degree between two intersections which can be computed using various functions. A Recursive Spatial Algorithm is implemented in order to find the optimal solution to the problem. This algorithm uses Lapacian eigenvectors to determine the partitions The reader is referred to Ma et al. (2009) for a detailed description of the approach. The authors apply the approach to a simple network, and to the Guangzhou network in China. The latter network consists of both urban roads and a freeway ring-road. The authors find that 17 partitions gives the best performance, these partitions are shown in Figure 4.13.

4.6.4 Discussion on partitioning approaches

Partitioning a network into sub-networks such that the dynamics of the sub-networks can be modeled using NFDs is not an easy task. An important issue in this area of research is that the selection of a

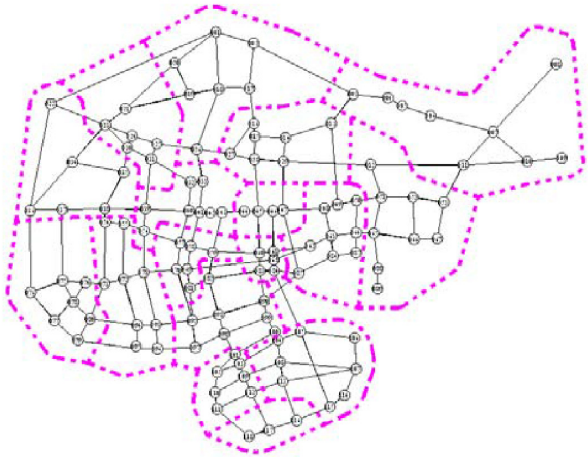

Figure 4.13: Result of Ma et al. (2009) of the partition of the Guangzhou, China road traffic network.

partition will affect the performance of the traffic flow model or traffic controller. So, when the partitions are made in order to control the flows between sub-networks, controller performance should be the main criterion for creating a partition. However, the choice of the control algorithm affects the NFD and the choice of the partition affects the controller performance. Thus, we end up with a chicken and egg problem which will not be easily solved.

Appendix A

Glossary

List of Acronyms

PI Proportional-Integral

MPC Model Predictive Control

MFD Macroscopic Fundamental Diagram

NFD Network Fundamental Diagram

Bibliography

Aboudolas, K. and Geroliminis, N. (2013). Feedback perimeter control for multi-region large-scale congested networks. In *Proceedings of the European Control Conference*, pages 3506–3511.

Ahn and Cassidy (2007). Freeway traffic oscillations and vehicle lane-change maneuvers. *Transportation and Traffic Theory*.

Bertini and Leal (2006). Empirical study of traffic features at a freeway lane drop. *Journal of Transportation Engineering*, 131:397–407.

Bliemer, M. (2007). Dynamic queuing and spillback in analytical multiclass dynamic network loading model. *Transportation Research Record: Journal of the Transportation Research Board*, (2029):14–21.

Cassidy, Daganzo, Jang, and Chung (2009). Spatiotemporal effects of segregating different vehicle classes on separate lanes. *Transportation and Traffic Theory*.

Daganzo and Cassidy (2006). Effects of high occupancy vehicle lanes on freeway congestion. *Transportation Research Part B: Methodological*, 42:861–872.

Daganzo, C. (2005a). Improving city mobility through gridlock control: an approach and some ideas. Report, UC Berkeley, Center of Excellence on Future Urban Transport.

Daganzo, C. (2005b). A variational formulation of kinematic waves: basic theory and complex boundary conditions. *Transportation Research Part B: Methodological*, 39:187–196.

Daganzo, C. (2005c). A variational formulation of kinematic waves: Solution methods. *Transportation Research Part B: Methodological*, 39:934–950.

Daganzo, C. (2007a). Urban gridlock: macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological*, 41(1):49–62.

Daganzo, C. and Geroliminis, N. (2008). An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological*, 42:771–781.

Daganzo, C. and Menendez, M. (2005). A variational formulation of kinematic waves: Bottleneck properties and examples. In *Transportation and Traffic Theory, Dynamics and Human Interaction. 16th International Symposium on Transportation and Traffic Theory 2005*.

Daganzo, C. F. (1995). The cell transmission model, part ii: network traffic. *Transportation Research Part B: Methodological*, 29(2):79–93.

Daganzo, C. F. (2005d). Improving city mobility through gridlock control: an approach and some ideas.

Daganzo, C. F. (2007b). Urban gridlock: macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological*, 41(1):49–62.

Etemadnia, H., Abdelghany, K., and Hassan, A. (2014). A network partitioning methodology for distributed traffic management applications. *Transportmetrica A: Transport Science*, 10(6):518–532.

Flötteröd, G. and Rohde, J. (2011). Operational macroscopic modeling of complex urban road intersections. *Transportation Research Part B: Methodological*, 45(6):903–922.

Geroliminis, N. and Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B-Methodological*, 42(9):759–770.

Geroliminis, N., Haddad, J., and Ramezani, M. (2013). Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach. *IEEE Transactions on Intelligent Transportation Systems*, 14(1):348–358.

Godfrey, J. (1969). The mechanism of a road network. *Traffic Engineering & Control*, 8(8).

Haddad, J., Ramezani, M., and Geroliminis, N. (2012). Model predictive perimeter control for urban areas with macroscopic fundamental diagrams. Proceedings of the American Control Conference, pages 5757–5762. Los Alamitos.

Haddad, J., Ramezani, M., and Geroliminis, N. (2013). Cooperative traffic control of a mixed network with two urban regions and a freeway. *Transportation Research Part B: Methodological*, 54:17–36.

Hajiahmadi, M., Knoop, V., De Schutter, B., and Hellendoorn, H. (2013). Optimal dynamic route guidance: A model predictive approach using the macroscopic fundamental diagram. In *Proceedings of the 16th IEEE Conference on Intelligent Transportation Systems*, pages 1022–1028. The Hague, The Netherlands.

Ji, Y. and Geroliminis, N. (2012). On the spatial partitioning of urban transportation networks. *Transportation Research Part B: Methodological*, 46(10):1639–1656.

Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., and Papageorgiou, M. (2012). Exploiting the fundamental diagram of urban networks for feedback-based gating. *Transportation Research Part B: Methodological*, 46(10):1393–1403.

Knoop, V. and Hoogendoorn, S. (2014). Network transmission model: a dynamic traffic model at nework level. In *Proceedings of the 93rd Annual Meeting of the Transportation Research Board*.

Knoop, V. L., Hoogendoorn, S., and Van Lint, J. (2013). The impact of traffic dynamics on macroscopic fundamental diagram. Transportation Research Board.

Laval and Daganzo (2006a). Lane-changing in traffic streams. *Transportation Research Part B: Methodological*, 40:251–264.

Laval, J. A. and Daganzo, C. F. (2006b). Lane-changing in traffic streams. *Transportation Research Part B: Methodological*, 40(3):251–264.

Lebacque, J.-P. (2005). First-order macroscopic traffic flow models: Intersection modeling, network modeling. In *Transportation and Traffic Theory. Flow, Dynamics and Human Interaction. 16th International Symposium on Transportation and Traffic Theory*.

Leclercq (2004). *Private communication*.

Leclercq, L. (2007). Hybrid approaches to the solutions of the "lighthill-whitham-richards" model. *Transportation Research Part B: Methodological*, 41:701–709.

Leclercq, L., Laval, J., and Chevallier, E. (2007). The lagrangian coordinates and what it means for first order traffic flow models. In Allsop, R., Bell, M., and Heydecker, B., editors, *17th International Symposium on Transportation and Traffic Theory*, pages 735–753. Elsevier.

Lighthill, M. and Whitham, G. (1995). On kinematic waves. i:flood movement in long rivers, ii: A theory of traffic flow on long crowded roads. In *Proceedings of Royal Society*, volume A229.

Luke, J. (1972). Mathematical models for landform evolution. *Journal of Geophysical Research*, 77:2460–2464.

Ma, Y.-Y., Chiu, Y.-C., and Yang, X.-g. (2009). Urban traffic signal control network automatic partitioning using laplacian eigenvectors. pages 1–5. IEEE.

Munoz and Daganzo (2002). Moving bottlenecks: A theory grounded on experimental observation. In *the 15th International Symposium on Transportation and Traffic Theory*, pages 441–462. PergamonElsevier, Oxford,UK.

Newell (1993). A simplified theory of kinematic waves in highway traffic, part 1: general theory. *Transportation Research Part B: Methodological*, 4:281–287.

Richards, P. (1956). Shock waves on the highway. *Opnr. Res.*, 4:42–51.

Smits, E.-S., Bliemer, M. C., Pel, A. J., and van Arem, B. (2015). A family of macroscopic node models. *Transportation Research Part B: Methodological*, 74:20–39.

Tampère, C. M., Corthout, R., Cattrysse, D., and Immers, L. H. (2011). A generic class of first order node models for dynamic macroscopic simulation of traffic flows. *Transportation Research Part B: Methodological*, 45(1):289–309.

Treiterer and Myers (1974). The hysteresis phenomenon in traffic flow. In *the 6th International Symposium on Transportation and Traffic Theory*.

Van der Gun, J. P., Pel, A. J., and Van Arem, B. (2015). A general activity-based methodology for simulating multimodal transportation networks during emergencies. In *ICEM 2015: 3rd International Conference on Evacuation Modeling and Management, Tainan, Taiwan, 1-3 June 2015*.

Yperman, I. (2007). The link transmission model for dynamic network loading.