Modelling and control of road traffic networks

Bertrand Haut

Membres du jury:

Georges Bastin (Promoteur)
Vincent Blondel
Guy Campion
Yacine Chitour
Jean-Patrick Lebacque
Paul Van Dooren

Remerciements

Je remercie mon promoteur, le Professeur G. Bastin, pour son encadrement efficace et ses critiques constructives qui m'ont permis de mener à bien ce travail de thèse.

Merci aux membres du jury pour leur relecture du manuscrit ainsi que pour la discussion fructueuse lors de la défense privée qui a activement contribué à la version finale de ce document.

Cette thèse n'aurait pas pu être possible sans le financement du Fonds National de la Recherche Scientifique. Que Madame Simoen trouve ici l'expression de mes remerciements les plus sincères.

Je voudrais également remercier Marie pour son soutien. Enfin, je ne voudrais pas oublier de remercier mes parents pour m'avoir permis de réaliser mes études ainsi que de l'attention qu'ils ont portée à ce travail, notamment lors de ces derniers mois.

Glossary

List of symbols

$A(u)$	a $n \times n$ matrix
$\mathcal{C}^1([a,b],\mathbb{R}^n)$	the space of continuously differentiable functions from $[a,b]$ to $\mathbb{R}^n$
$f(\rho)$	flow in function of the surrounding density
$G(s)$	an input–output transfer function
L_i	length of road i
λ_i	i^{th} eigenvalue
∇	the gradient operator
Ω	an open set limiting the possible values of the state variable
∂_x	partial derivative with respect to x
∂D	boundary of D
q	flow
ρ	density: number of vehicle per time unit
$\rho^{\max}$	maximal density on a road
$\rho(A)$	spectral radius of matrix A
$r \bullet f(u)$	directional derivative of f in the direction r
$\mathbb{R}$	set of real
$R(\rho)$	receiving capacity
$S(\rho)$	sending capacity
σ	critical density for the LWR model $\sigma = \arg\max_\rho f(\rho)$
σ_K	critical density for the AR model associated to the curve Υ_K
$\overset{\circ}{D}$	interior of D
t	time
$u(x,t)$	a state variable
$\bar{u}(x)$	initial condition of the state variable
u_l	for a Riemann problem: the left initial state
u_r	for a Riemann problem: the right initial state
$u_{i,0}$	for a Riemann problem: initial state on road i
$\bar{u}_i$	for a Riemann problem: new state on the road i near a junction after the first interaction

$U(x,t)$	a traffic state variable in $\mathbb{R}^2$: $U(x,t) = (\rho(x,t), v(x,t))$
Υ	a curve in the $(\rho, \rho v)$ plane corresponding to a fixed value of $v + p(\rho)$
Υ_K	a curve in the $(\rho, \rho v)$ plane corresponding to a fixed value of $v + p(\rho) = K$
v	speed of the drivers
$V(\rho)$	preferential speed in function of the surrounding density
$v^{\max}$	maximal speed on a road
x	position on the road
ξ_i	i^{th} Riemann invariant

Acronyms and abbreviations

PDE	Partial Differential Equation
LWR	the Lighthill–Whitham–Richards single order model
AR	the Aw & Rascle second order model
TTS	Total Time Spent
OD	Origin–Destination
UO	User–Optimal
SO	System–Optimal
BV	the space of functions with bounded variations

Contents

Chapter 1

Introduction

With the constant growth of the number of vehicles, the congestion problems on highways are becoming more and more worrying. This is particularly obvious near big cities at morning and evening rush hours. These problems are not ready to disappear; indeed it is estimated that the traffic has a considerable annual growth rate from six to twenty percent in function of the country. These congestion problems have not only a high cost but also various negative effects on safety, environment and the driver's nerves.

To limit these problems, it is not always possible to build new roads, especially near big cities. It is even sometimes counterproductive. In addition to develop the public transports, it is thus necessary to increase the efficiency and safety of the existing traffic networks. To perform this increase, the notion of Intelligent Traffic Systems (ITS), sometimes called Intelligent Transportation System, was introduced. ITS encompass a broad range of communications-based information and electronic technologies. With ITS, the highways are no more passive but interact with the drivers in order to increase the global efficiency of the network.

ITS includes classical applications such as cameras to monitor the traffic flow, which allows to give informations over the traffic state on the radio or on the Internet. Combined with variable message signs (VMS), it is possible to warn directly the drivers of accidents or other delays. Moreover, these variable messages can also be used to adjust the maximum legal speed in function of the number of vehicles or the current weather.

Another classical technology used in ITS is the use of traffic lights to perform ramp metering. This allows, for example, to limit the flow entering on a highway in order to prevent the appearance of a traffic jam. More complex applications can also be considered. One can think over adjusting the toll costs on some road in order to stimulate the drivers to choose an alternative path. Correctly implemented, it would allow to balance the traffic flow on the different paths. A last interesting technology is the implementation of reversible flow lanes. Using lane control signs, the direction of a lane could be reversed allowing travel in the peak direction during rush hours or for special events.

Some of these ITS are already applied in various countries. The traffic state on the ring of Brussels is available on the Internet[1]. In the Netherlands, variable speed regulation systems are commonly used. The speed limit is lowered if a start of congestion is measured. Ramp metering is common in some places of North America such as Los Angeles, Phoenix and Minneapolis-St. Paul metropolitan areas.

If the highways are equipped with sensors allowing to perform measurements on the traffic states, one can think over some feedback strategies where some actions are automatically undertaken in function of the current traffic state. In order to develop such feedback strategies, a model describing the evolution of the traffic state is needed. This model should be able to represent as much as possible the phenomena observed on highways while being simple enough to be analysed mathematically and used for control purposes. A part of this work is thus dedicated to the modelling. It means the achievement of a set of variables sufficient to describe the traffic state and a set of equations allowing the prediction of the evolution of the variables. There exist mainly three families of models:

The microscopic models take into account the position and speed of each vehicle. The evolution of each vehicle is described by an ordinary differential equation expressing the behaviour of the drivers under the action of the surrounding vehicles. Due to the modelling at very small scale, these models are ideally suitable for urban traffic where there is a lot of interactions with the environment (traffic lights, intersection...). However, the simulation of a large road network using such models is time consuming. These models are thus rarely used for highways networks. A survey can be found in [21].

The kinetic models, sometimes called statistical models, introduce a distribution function in space and velocity instead of a description of individual cars. These models can be derived from microscopic considerations and can be seen as an intermediate modelling between microscopic and macroscopic modelling. They are based on Boltzmann type kinetic equations and already allow faster computation than the microscopic models. These models started with the work of Prigogine [55].

The macroscopic models take an approach similar to hydrodynamics. The traffic is viewed as a liquid flowing in tubes representing the roads. This is done by introducing a density and a speed function. These two functions depend on the position on the road and on the time. They represent respectively the number of vehicles per space unit and the average speed. These models are thus only valid at a sufficiently large space and time scale (50 to 100 meters). Such modelling on a large scale allows very fast computation. This is the reason why such models are frequently

[1] http://www.verkeerscentrum.be/verkeersinfo/camera_overzicht_brussel

used on large highway networks. On the other side, these models are not designed to easily take into account traffic lights, intersections and very small roads. These models are thus rarely used for urban traffic.

The first model of this type was introduced by Lighthill, Whitham and Richards (LWR) [49, 60]. Various other more complex models [54, 60, 40] have been developed since then. Daganzo in his "Requiem for second-order fluid approximations of traffic flow" [10] pointed out some important limitations of these models. But recently some new models [2] have been developed which overcome these limitations.

Two main issues will be considered in this thesis: modelling and control.

Modelling aspects

In this work, the macroscopic models are used. It means that the modelling and the traffic control are mainly focused on highway traffic and not urban traffic. Two models are presented and extended.

The first one is the *single order* LWR model [49, 60]. The term single order is used to point out that a single partial differential equation (PDE) is used to describe the evolution of the traffic state. This model is presented and extended in Chapter 3. The contributions are mainly focused on the junction modelling and particularly on the representation of the *capacity drop* phenomenon. This critical phenomenon represents the fact that the outflow of a junction can be significantly lower than the maximum achievable flow at the same location. It can easily be understood in the case of a merge junction where two roads merge in one: if there are too many vehicles trying to access the same road, there is some kind of mutual hindrance between the drivers which results in an outgoing flow lower than the optimal possible flow. This modification of the LWR model in order to represent the capacity drop phenomenon has been presented in [26]. A numerical simulator using this single order model has also been developed. The parameters used by this simulator are calibrated using experimental measurements.

The second model is the Aw & Rascle (AR) model [2]. This is a second order model since two PDE's are used to represent the traffic evolution. This model is presented and extended in Chapter 5. The two main contributions made in this thesis are a stability analysis and a junction model. The core of the junction modelling is a new assumption describing the drivers' behaviour at the junction. This new assumption and the properties of the corresponding junction model have been presented in [25, 23]. The properties analysed are the capacity drop phenomenon which is naturally represented by this junction model and the stability of equilibria which allows to prove the existence of solutions for a large class of initial conditions.

Each of these two models has advantages and disadvantages. The first model is simpler and allows very fast computation while the second model is able to represent more complex phenomena.

Control aspects

Chapter 4 of this thesis deals with the control aspects and can mainly be divided in three parts: static optimisations, a routing strategy and a linear feedback ramp metering method. Each of these control methods is illustrated using the single order LWR model.

In the first part, a realistic cost function is developed to represent the global efficiency of a road network. This function is derived from data provided by economists and includes the costs related to the spent time, the generated pollution and the risk of accident. On the basis of this cost function various optimisations can be performed to analyse the best flow repartitions, the influence of some speed limitations or the construction of a new road. This work has led to [29].

In the second part, a routing strategy is developed. Usually there are different paths available for a driver to reach a particular destination from a particular origin. It can be assumed that it is possible, for a network operator, to influence the fraction of the drivers taking one of those paths, for example by presenting some informations about the time needed on each path. The problem of "optimal routing" consists for the operator to adjust the "splitting rate" at the bifurcation nodes in order to optimise a cost criterion representing the efficiency of the network. The routing strategy developed in this thesis provides a rule to adapt the splitting rate and has been presented in [24].

In the third part, a ramp metering strategy using linear feedback is developed. The idea behind this control method is to stabilise the system which prevents the appearance of a traffic jam. It is shown that this stabilisation problem can be rewritten as a spectral radius problem. A part of this thesis is thus dedicated to the resolution of this algebraic problem and its application to the ramp metering case. This result has been presented in [28, 27].

Publication list

Journal articles

[25] B. Haut, G. Bastin, A second order model of road junctions in fluid models of traffic networks, AIMS Journal on Networks and Heterogeneous Media (NHM), Vol. 2(2), pp. 227–253, 2007.

[27] B. Haut, G. Bastin, P. Van Dooren, Maximal stability region of a perturbed nonnegative matrix, submitted for publication.

Conference proceedings

[26] B. Haut, G. Bastin, Y. Chitour, A macroscopic traffic model for road networks with a representation of the capacity drop phenomenon at the

junction, *in Proc. of* 16th IFAC World Congress, Prague, Czech Republic, July 4–8, 2005.

[28] B. Haut, G. Bastin, P. Van Dooren, Maximal nonnegative perturbation of a nonnegative matrix, *in Proc. of* 2nd International Symposium on Positive Systems POSTA 06, Grenoble, France, 30 August - 2 September 2006, Lecture Notes in Control and Information Sciences, Springer Verlag, Vol. 341, pp.201-208.

[24] B. Haut, G. Bastin, A System-Optimal Routing Policy For Road Traffic Networks, *in Proc. of* 11th IFAC Symposium on Control in Transportation Systems, Delft, The Netherland, August 29–31, 2006.

[23] B. Haut, G. Bastin, A second order model for road traffic networks, *in Proc. of* ITSC'05 – 8th International IEEE Conference on Intelligent Transportation Systems, September 2005, pp. 178–184.

Technical report

[29] B. Haut, H. Cote, G. Bastin, G. Campion, DWTC-CP/40 Final report: Sustainability effects of traffic management systems, 2005.

Chapter 2

Mathematical preliminaries

In this thesis, hyperbolic systems will be used to develop fluid models describing the evolution of road traffic. In this chapter, the main properties of hyperbolic systems are presented. For a complete study we refer to [6] from which most of the results come from or to [56] and [60].

2.1 Quasilinear hyperbolic systems

A *quasilinear system* of n equations in one space variable takes the form

$$\partial_t u + A(u)\partial_x u = 0 \tag{2.1}$$

where ∂_t (resp. ∂_x) denotes the partial derivative with respect to t (resp. x) and $A(u)$ is a $n \times n$ matrix. The state variable $u(x,t)$ belongs to an open set $\Omega \subseteq \mathbb{R}^n$, the space variable x belongs to $\mathbb{R}$ and the time variable t to $\mathbb{R}^+$.

Remark 2.1.1 (Conservation laws)**.** *Such quasilinear systems appear naturally when considering problems where some quantities are conserved. A particular example is a model describing the evolution of the traffic; in that case the total number of vehicles is conserved. Such systems can be written as follows:*

$$\partial_t u + \partial_x f(u) = 0 \tag{2.2}$$

where $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ is called the flux *and $u \in \Omega$ is called the density and represents the conserved quantity. Indeed, integrating* (2.2) *over the space interval $[a,b]$ and assuming $u \in \mathcal{C}^1$, we have*

$$\begin{aligned}
\int_a^b \partial_t u(x,t)dx &= -\int_a^b \partial_x f(u(x,t))dx \\
\partial_t \int_a^b u(x,t)dx &= f(u(a,t)) - f(u(b,t)) \\
&= \textit{[inflow at a] - [inflow at b]} .
\end{aligned}$$

This means that the total variation of the quantity u over the space interval $[a, b]$ is only due to the inflow at the positions a and b. No quantity of u is created or destroyed between a and b. If the function $f(u)$ is regular enough, (2.1) *can be seen as a reformulation of* (2.2) *by defining*

$$A(u) = \nabla_u f(u).$$

Definition 2.1.2 (Classical solution)**.** *A classical solution of* (2.1) *is a continuously differentiable function $u = u(x, t)$ which satisfies* (2.1) *at every point of its domain. If an initial condition $\bar{u}(x)$ is given, u should also satisfies $u(x, 0) = \bar{u}(x)$.*

Even for "nice" functions $A(u)$ and $\bar{u}(x)$ (for example belonging to $\mathcal{C}^1$), it is possible that no classical solution to (2.1) exists. In this case, some solutions may exist if the time variable t is restricted to a bounded interval $[0, T]$.

Definition 2.1.3 (Hyperbolicity)**.** *The system* (2.1) *is said* hyperbolic *if for every $u \in \Omega$ the matrix $A(u)$ has n real distinct eigenvalues $\lambda_1 \leq \ldots \leq \lambda_n$. If the inequalities are strict the system is said* strictly hyperbolic*.*

If the system presents the property of strict hyperbolicity, it is sometimes possible to prove the existence of a classical solution on a sub-domain.

Theorem 2.1.4 (Local existence)**.** *For the system* (2.1) *with the initial conditions*

$$u(x, 0) = \bar{u}(x) \qquad x \in [a, b],$$

if the system is strictly hyperbolic and the functions $A(u)$ and $\bar{u}(x)$ are continuously differentiable then there exist constants C, $\delta > 0$ and a function $u(x, t)$ which is the unique classical solution on the domain

$$\mathcal{D} = \{(x, t) \mid a + Ct \leq x \leq b - Ct,\ t \in [0, \delta]\}.$$

2.2 Riemann invariants

For some hyperbolic systems, there exists a change of variables $u(x, t) \rightarrow \boldsymbol{\xi}(u(x, t))$ such that (2.1) can be transformed into a system of coupled transport equations.

A *Riemann invariant* is a function $\xi_i(u) : \Omega \mapsto \mathbb{R}$ which, according to the original system (2.1), satisfies

$$\partial_t \xi_i(u(x, t)) + \lambda_i(u(x, t)) \partial_x \xi_i(u(x, t)) = 0.$$

It can be seen that these equations express the total derivative of ξ_i ($d\xi_i/dt$) on a particular curve which is defined by the following equation

$$\frac{dx}{dt} = \lambda_i(u(x, t)).$$

This curve is called a *characteristic curve.* Since the total derivate of ξ_i is equal to zero on this curve, the value of ξ_i remains constant along this characteristic curve. The slope λ_i of the characteristic curve in the (x,t) plane are given by the eigenvalues of $A(u)$.

If there exist n independent Riemann invariants $\xi_1, \dots, \xi_n$, it is possible to diagonalise (2.1) using the change of variables defined by the Riemann invariants:

$$\partial_t \boldsymbol{\xi} + \Lambda(\boldsymbol{\xi})\partial_x \boldsymbol{\xi} = 0$$

where

$$\boldsymbol{\xi} = \begin{pmatrix} \xi_1(u) \\ \vdots \\ \xi_n(u) \end{pmatrix} \quad \text{and} \quad \Lambda(\boldsymbol{\xi}) = \begin{pmatrix} \lambda_1(u(\boldsymbol{\xi})) & & \\ & \ddots & \\ & & \lambda_n(u(\boldsymbol{\xi})) \end{pmatrix}.$$

Remark 2.2.1. *With this change of variable, it is obvious that the growth rate of the influence area of an "event" occurring at a particular x is bounded by* $\min_i \lambda_i$ *and* $\max_i \lambda_i$*. In particular, it implies that the value $u(x^*, t^*)$ does not depend of $u(x,t)$ with*

$$x - x^* < C(t - t^*)$$

where C is the maximal speed of the characteristic curves

$$C = \max_{u \in \Omega, i} \lambda_i(u).$$

See Figure 2.1 for the illustration of this area.

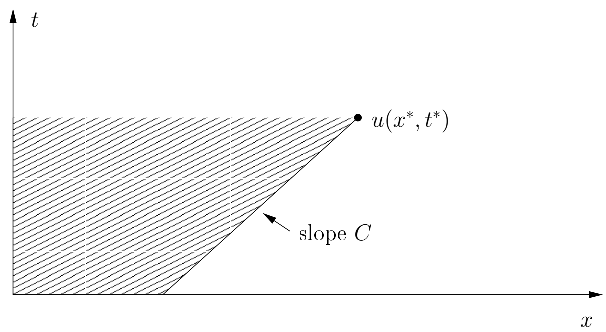


Figure 2.1: The hatched area has no influence on the value of $u(x^*, t^*)$.

2.3 Stability

The stability of a system which can be expressed in terms of his Riemann invariants can be analysed using the extension of the Li Ta-sien's theorem (see [48]) proposed in [14]. This theorem can also be used to guarantee the existence of a classical solution.

Let $0 \leq m \leq n$ be two integers. Let $\Lambda : \mathbb{R}^n \mapsto \mathbb{R}^{n\times n}$ be a continuously differentiable function in a neighbourhood of $0 \in \mathbb{R}^n$ such that

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$$

with $\lambda_i(0) < 0, \forall i \in \{1, \ldots, m\}$ and $\lambda_i(0) > 0, \forall i \in \{m+1, \ldots, n\}$. For example the λ_i can be the eigenvalues of the matrix $A(u)$.

Let $\Lambda_- : \mathbb{R}^m \mapsto \mathbb{R}^{m\times m}$ and $\Lambda_+ : \mathbb{R}^{n-m} \mapsto \mathbb{R}^{(n-m)\times(n-m)}$ be the two functions defined by

$$\Lambda_- = \mathrm{diag}(\lambda_1, \ldots, \lambda_m), \qquad \Lambda_+ = \mathrm{diag}(\lambda_{m+1}, \ldots, \lambda_n).$$

For all $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)^T \in \mathbb{R}^n$, let $\boldsymbol{\xi}_- = (\xi_1, \ldots, \xi_m)^T \in \mathbb{R}^m$ and $\boldsymbol{\xi}_+ = (\xi_{m+1}, \ldots, \xi_n)^T \in \mathbb{R}^{n-m}$.

For $x \in [a, b]$ and $t \in \mathbb{R}$, the following hyperbolic system is considered

$$\partial_t \boldsymbol{\xi} + \Lambda(\boldsymbol{\xi})\partial_x \boldsymbol{\xi} = 0, \tag{2.3}$$

together with the boundary condition

$$\begin{pmatrix} \boldsymbol{\xi}_-(b,t) \\ \boldsymbol{\xi}_+(a,t) \end{pmatrix} = \mathbf{g} \begin{pmatrix} \boldsymbol{\xi}_-(a,t) \\ \boldsymbol{\xi}_+(b,t) \end{pmatrix} \tag{2.4}$$

where $\mathbf{g} : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a continuously differentiable function in a neighbourhood of $0 \in \mathbb{R}^n$ satisfying $\mathbf{g}(0) = 0$.

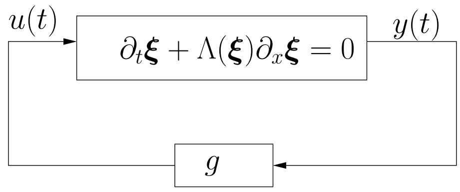


Figure 2.2: The outputs of the system are the Riemann invariants leaving the domain $y(t) = (\boldsymbol{\xi}_-(a,t), \boldsymbol{\xi}_+(b,t))^T$ and the inputs are the Riemann invariants entering the domain $u(t) = (\boldsymbol{\xi}_-(b,t), \boldsymbol{\xi}_+(a,t))^T$.

For example, one can think of the original system (2.1) where the change of variables defined by the Riemann invariants has been applied. The function g can be seen as a feedback law which connects the output to the input (see Fig. 2.2). Such systems are met when ramp metering strategies using linear feedback are considered (see Section 4.4). If the initial state satisfies a particular *compatibility* condition then a sufficient stability condition can be derived.

Definition 2.3.1. *The function* $\boldsymbol{\xi}^\# \in \mathcal{C}^1([a,b];\mathbb{R}^n)$ *satisfies the compatibility condition* (C) *if*

$$\begin{pmatrix} \boldsymbol{\xi}_-^\#(b) \\ \boldsymbol{\xi}_+^\#(a) \end{pmatrix} = \boldsymbol{g} \begin{pmatrix} \boldsymbol{\xi}_-^\#(a) \\ \boldsymbol{\xi}_+^\#(b) \end{pmatrix}$$

$$\begin{pmatrix} \Lambda_-(\boldsymbol{\xi}^\#(b))\frac{\partial \boldsymbol{\xi}_-^\#}{\partial x}(b) \\ \Lambda_+(\boldsymbol{\xi}^\#(a))\frac{\partial \boldsymbol{\xi}_+^\#}{\partial x}(a) \end{pmatrix} = \nabla \boldsymbol{g} \begin{pmatrix} \boldsymbol{\xi}_-^\#(a) \\ \boldsymbol{\xi}_+^\#(b) \end{pmatrix} \begin{pmatrix} \Lambda_-(\boldsymbol{\xi}^\#(a))\frac{\partial \boldsymbol{\xi}_-^\#}{\partial x}(a) \\ \Lambda_+(\boldsymbol{\xi}^\#(b))\frac{\partial \boldsymbol{\xi}_+^\#}{\partial x}(b) \end{pmatrix}.$$

Theorem 2.3.2. *If*

$$\rho(|\nabla \boldsymbol{g}(0)|) < 1, \tag{2.5}$$

where ρ denotes the spectral radius, then there exist $\epsilon > 0, \mu > 0$ and $C > 0$ such that, for every $\boldsymbol{\xi}^\# \in \mathcal{C}^1([a,b];\mathbb{R}^n)$ satisfying (C) *and such that*

$$\left\|\boldsymbol{\xi}^\#\right\|_{C^1(a,b)} \leq \epsilon,$$

there exists one and only one function $\boldsymbol{\xi} \in C^1([a,b] \times [0,+\infty);\mathbb{R}^n)$ satisfying (2.3)*,* (2.4) *and*

$$\boldsymbol{\xi}(x,0) = \boldsymbol{\xi}^\#(x) \qquad \forall x \in [a,b].$$

Moreover, this function $\boldsymbol{\xi}$ satisfies

$$\|\boldsymbol{\xi}(\cdot,t)\|_{\mathcal{C}^1(a,b)} \leq Ce^{-\mu t}\left\|\boldsymbol{\xi}^\#\right\|_{\mathcal{C}^1(a,b)} \qquad \forall t \geq 0.$$

The $\mathcal{C}^k$ norm used in the previous theorem is defined as

$$\|u(\cdot,t)\|_{C^k([a,b])} = \sum_{n=0}^{k} \sup_{x\in[a,b]} \left|u^{(n)}(x,t)\right|.$$

If the system represented in Fig 2.2 is considered, (2.5) expresses that the linearisation of the function g connecting the output $y(t)$ to the input $u(t)$ has a gain less than one. This condition can thus be viewed as a "small gain" condition. Since the system (2.3) is a set of delay equations, it has an unit gain. The total gain of the closed loop system represented in Fig 2.2 is thus less than one and it is natural to expect convergence to an equilibrium.

2.4 Discontinuous solutions

As already mentioned, a classical solution $u(x,t)$ does not always exist on the domain $\mathbb{R}\times\mathbb{R}^+$ even for smooth initial condition $\bar{u}(x)$. For example, considering the simple system

$$\partial_t u + u\left(1 - \frac{u}{2}\right)\partial_x u = 0$$

with $u(x,t) \in \mathbb{R}$ and the initial condition

$$\bar{u}(x) = 1 + 0.5 \text{ atan}(x-5),$$

different characteristic curves carrying different values of u meet on the same point (x^*,t^*) of the domain (see Fig. 2.3). It means that

$$\lim_{\substack{t\to t^* \\ <}} \partial_x u(x^*,t) = +\infty$$

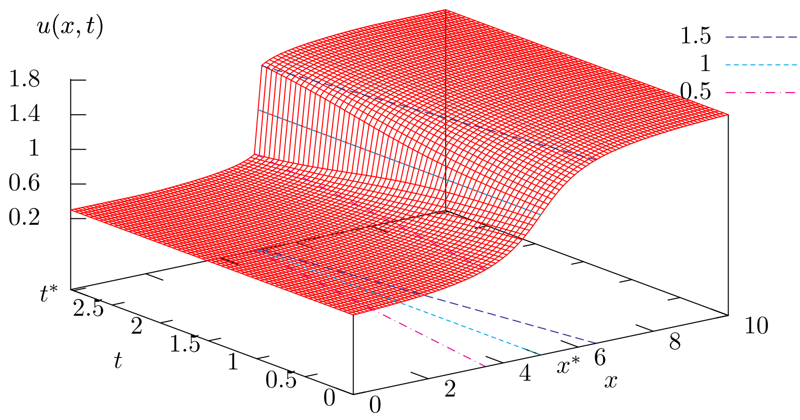


Figure 2.3: A blowup of $\partial_x u$ in a finite time. The three isolines corresponding to three different values of u are converging to the same (x, t) in a finite time.

and no classical solution exists on the domain $\mathbb{R} \times \mathbb{R}^+$.

Instead of reducing the domain of the solution, it is possible to extend the concept of solution to allow discontinuities. For example, one could look for functions $u(x, t)$ which are piecewise Lipschitz continuous in the x–t plane with jumps around a finite number of curves $x = \gamma_j(t),\ j = 1, \ldots, N$. Such discontinuities are admissible in the context of traffic. One can think of the tail of a traffic jam where the density increases very quickly from a low to a high density. On a macroscopic scale this can be approximated by a discontinuity. Another example is near a red traffic light where there is a real discontinuity: upstream the road is crowded and downstream the road is empty.

2.4.1 Weak solutions

Let $f : \Omega \mapsto \mathbb{R}^n$ be a smooth vector field and consider the following system of conservation laws

$$\partial_t u + \partial_x f(u) = 0. \tag{2.6}$$

It is possible to define the concept of a *distributional solution* of (2.6) without the expression of the partial derivative of u. This allows the presence of discontinuities.

Definition 2.4.1. *A measurable function $u(x, t)$ from an open set $\Omega \subseteq \mathbb{R} \times \mathbb{R}$ into $\mathbb{R}^n$ is a* distributional solution *of* (2.6) *if for every $\mathcal{C}^1$ function $\Phi : \Omega \mapsto \mathbb{R}^n$ with compact support, one has*

$$\iint_\Omega \{u \partial_t \Phi + f(u) \partial_x \Phi\} dx dt = 0.$$

The distributional solution u is not necessarily continuous but u and $f(u)$ should be locally integrable in Ω. This definition can be extended to take into account the existence of initial conditions.

Definition 2.4.2. *Given an initial condition*

$$u(x,0) = \bar{u}(x), \tag{2.7}$$

with $\bar{u} \in \mathbf{L}^1_{loc}(\mathbb{R};\mathbb{R}^n)$*, a function* $u : [0,T] \times \mathbb{R} \mapsto \mathbb{R}^n$ *is a* distributional solution *of the Cauchy problem* (2.6)–(2.7) *if*

$$\int_0^T \int_{-\infty}^{+\infty} \{u\partial_t\Phi + f(u)\partial_x\Phi\} dxdt + \int_{-\infty}^{+\infty} \bar{u}(x)\Phi(x,0)dx = 0$$

for every $\mathcal{C}^1$ *function* Φ *with compact support contained in the set* $]-\infty, T[\times\mathbb{R}$*.*

There is also a stronger concept of discontinuous solution requiring the continuity of u as a function of time with values into $\mathbf{L}^1_{\text{loc}}(\mathbb{R})$.

Definition 2.4.3 (Weak solution)**.** *A function* $u : [0,T] \times \mathbb{R} \mapsto \mathbb{R}^n$ *is a* weak solution *of the Cauchy problem* (2.6)–(2.7) *if* u *is continuous as a function from* $[0,T]$ *into* $\mathbf{L}^1_{loc}$ *, the initial condition* (2.7) *holds and the restriction of* u *to the open strip* $]0,T[\times\mathbb{R}$ *is a distributional solution of* (2.6)*.*

Every weak solution is a distributional solution but the converse is clearly false.

2.4.2 Rankine–Hugoniot condition

It is possible to obtain a condition that $u(x,t)$ must satisfy around a discontinuity to be a distributional solution of (2.6). Consider first the simplest possible discontinuous solution

$$U(x,t) = \begin{cases} u_r & \text{if } x > \lambda t, \\ u_l & \text{if } x \le \lambda t, \end{cases} \tag{2.8}$$

for some $u_l, u_r \in \mathbb{R}^n, \lambda \in \mathbb{R}$.

Lemma 2.4.4. *The function* U (2.8) *is a solution of* (2.6) *iff*

$$\lambda(u_r - u_l) = f(u_r) - f(u_l). \tag{2.9}$$

This condition (2.9) is called the *Rankine-Hugoniot* condition and relates the two constant states to the speed λ of the discontinuity. This condition ensures that the total quantity of u is conserved even around the discontinuity. It is possible to derive the same condition for a more general solution $u(x,t)$ of (2.6).

Theorem 2.4.5. *Let u be a bounded distributional solution of* (2.6) *and fix a point* (ξ, τ). *For some states u_l, u_r and a speed λ, and defining U as in* (2.8), *assume that*

$$\lim_{r\to 0} \frac{1}{r^2} \int_{-r}^{r} \int_{-r}^{r} |u(\xi + x, \tau + t) - U(x,t)|\, dxdt = 0.$$

Then the Rankine-Hugoniot condition (2.9) *holds.*

It means that if (2.8) is a good approximation of $u(x,t)$ around (ξ, τ) then the Rankine-Hugoniot condition holds.

2.4.3 Admissibility conditions

For a given initial state, it is sometimes possible to find multiple weak solutions. In order to achieve uniqueness and continuous dependence on the initial data, additional "admissibility conditions" must be added. Various conditions such as *vanishing viscosity*, *entropy inequality* or *Lax condition* are possible. In this work, we use only the Lax condition.

Lax condition. A weak solution $u(x,t)$ of (2.6) is admissible if, at every point (τ, ξ) of approximate jump, the left and right states u_l, u_r and the speed λ of the jump satisfy

$$\lambda_i(u_l) \geq \lambda \geq \lambda_i(u_r) \tag{2.10}$$

for some $i \in \{1, \ldots, n\}$ where $\lambda_i(u)$ is the i^{th} eigenvalue of $Df(u)$, the $n \times n$ Jacobian matrix of f at u.

There is a simple geometric interpretation of the Lax condition. The characteristic curves associated to the eigenvalue i (i.e. with speed equal to $\lambda_i(u)$) must, on both sides of the discontinuity, run towards the line of discontinuity.

The choice of this condition is motivated by the fact that, for a system of one or two conservation laws, the solutions violating the Lax's condition across shocks are unstable. More precisely (see [9]), consider the system (2.6) with initial condition

$$\bar{u}(x) = \begin{cases} u_l & \text{if } x < 0 \\ u_r & \text{if } x > 0 \end{cases}$$

such that (2.8) is a solution which satisfy the Rankine–Hugoniot condition. If the Lax's condition is not satisfied then for certain smooth initial conditions $\bar{u}_n(x)$ converging to $\bar{u}(x)$, (2.6) have smooth solutions which are defined for all (x,t), $t \geq 0$, and which converge not to $U(x,t)$ but to a Lipschitz continuous solution of (2.6). In this work, only systems of one or two conservation laws are used. In this case it is natural to analyse, among the possible solutions, the stable ones since only these solutions appear in practice.

2.4.4 Riemann problem

The Riemann problem is the simplest problem involving a discontinuity in the initial condition:

$$\begin{aligned}\partial_t u + \partial_x f(u) &= 0 \\ u(x,0) &= \begin{cases} u_l & \text{if } x < 0 \\ u_r & \text{if } x > 0 \end{cases}\end{aligned} \tag{2.11}$$

where $u_l, u_r \in \Omega$. This problem is important since it is one of the few which admits analytical solutions. Moreover if a problem with complex initial condition is considered, the solution can be approximated through the resolution of multiple Riemann problems. An example of a Riemann problem is when a traffic light switches from red to green. Initially there is a discontinuity in the density which evolves when the first drivers begin to move.

The Jacobian $n \times n$ matrix of $f(u)$ is denoted $A(u) = Df(u)$. The system (2.6) is said to be *strictly hyperbolic* if the matrix $A(u)$ has n real distinct eigenvalues $\lambda_1 < \ldots < \lambda_n$. For such system, it is possible to find independent right and left normalised eigenvectors of $A(u)$ $\{r_1, \ldots, r_n\}, \{l_1, \ldots, l_n\}$:

$$\begin{aligned} l_i(u)A(u) &= \lambda_i(u) l_i(u) \\ A(u) r_i(u) &= \lambda_i(u) r_i(u) \\ |r_i| &= |l_i| = 1 \\ l_i \cdot r_j &= \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}. \end{aligned}$$

The directional derivative of a function g in the direction r_i is denoted as follow:

$$r_i \bullet g(u) = \lim_{\epsilon \to 0} \frac{g(u + \epsilon r_i) - g(u)}{\epsilon}.$$

The i-characteristic can be *genuinely non-linear* or *linearly degenerate*. This property influences the solution of the Riemann problem.

Definition 2.4.6. *For $i \in \{1, \ldots, n\}$, the i-characteristic field is said* genuinely non-linear *if*

$$r_i \bullet \lambda_i(u) \neq 0 \qquad \textit{for all } u \in \Omega.$$

If instead

$$r_i \bullet \lambda_i(u) = 0 \qquad \textit{for all } u \in \Omega,$$

the i-characteristic field is said to be linearly degenerate.

If the i-characteristic is genuinely non-linear, after having possibly changed the sign of r_i, it can be assumed that

$$r_i \bullet \lambda_i(u) > 0 \qquad \text{for all } u \in \Omega.$$

If each characteristic of (2.6) are either genuinely non-linear or linearly degenerate, it is possible to find an analytical solution for some of the Riemann problems (2.6)–(2.11). Two special cases are first considered: the *rarefaction wave* and the *shock wave*. These solutions are self-similar (i.e. $u(x,t) = \psi(x/t)$). After that, the general case is taken in consideration as a combination of these special cases.

The rarefaction wave

The first considered special case is the *rarefaction wave*. This solution of the Riemann problem (2.6)–(2.11) is a smooth self-similar function.

Denote by $\sigma \mapsto R_i(\sigma)(u_0)$ the parametrised integral curve of the eigenvector r_i through the point u_0, i.e. the solution of the Cauchy problem

$$\frac{du}{dt} = r_i(u(t)), \qquad u(0) = u_0.$$

Theorem 2.4.7. *If the initial states u_l, u_r are such that*

$$u_r = R_i(\bar{\sigma})(u_l)$$

for some $\bar{\sigma} \geq 0$ and some i, then the following piecewise smooth function

$$u(x,t) = \begin{cases} u_l & \text{if } x/t < \lambda_i(u_l) \\ u_r & \text{if } x/t > \lambda_i(u_r) \\ R_i(\sigma)(u_l) & \text{if } x/t \in [\lambda_i(u_l), \lambda_i(u_r)],\ x/t = \lambda_i(R_i(\sigma)(u_l)) \end{cases} \tag{2.12}$$

is a weak solution of (2.6)–(2.11). *This particular solution is called a* rarefaction wave.

For example, consider the following quasi-linear system

$$\begin{aligned} \partial_t \begin{pmatrix} u^1 \\ u^2 \end{pmatrix} \quad &+ \quad \begin{pmatrix} u^2 & u^1 \\ 0 & u^2 - u^1 p'(u^1) \end{pmatrix} \partial_x \begin{pmatrix} u^1 \\ u^2 \end{pmatrix} = 0 \\ u(x,0) \quad &= \quad \begin{cases} (1,1)^T & \text{if } x < 0 \\ (0.2, 2.6)^T & \text{if } x > 0 \end{cases} \end{aligned} \tag{2.13}$$

with $p(u^1) = \ln(u^1)$. This system can be derived from the following system of conservation laws

$$\partial_t \begin{pmatrix} u^1 \\ y \end{pmatrix} + \partial_x \begin{pmatrix} \left(\frac{y}{u^1} - p(u^1)\right) u^1 \\ \left(\frac{y}{u^1} - p(u^1)\right) y \end{pmatrix} = 0 \tag{2.14}$$

with $y = u^1(u^2 + p(u^1))$. The eigenvector r_1 associated to the first eigenvalue of the matrix in (2.13) is

$$r_1 = \begin{pmatrix} -1 \\ p'(u^1) \end{pmatrix}.$$

Since the function $u(t) : \mathbb{R} \mapsto \mathbb{R}^2$ such that

$$p(u^1(t)) + u^2(t) = C^{\mathrm{st}} \tag{2.15}$$

is a solution of

$$\frac{du(t)}{dt} = r_1(u(t)),$$

the parametrised integral curve $R_1(\sigma)(u_l)$ is obtained by selecting the function (2.15) passing through u_l (see Fig. 2.4(a)). Since the state u_r belongs to this curve, the solution of the Riemann problem consists of a single relaxation wave (2.12) between the two initial state u_l and u_r (see Fig. 2.4(b)).

This rarefaction wave corresponds to the situation at a traffic light switching from red to green. The first drivers start up and accelerate progressively allowing the next drivers to start up.

The shock wave

The second special case considered is the *shock wave.* This solution of the Riemann problem (2.6)–(2.11) is a discontinuity which travels at a constant speed, i.e. a solution of the form

$$u(x,t) = \begin{cases} u_l & \text{if } x < \lambda t, \\ u_r & \text{if } x > \lambda t, \end{cases}$$

for some λ. Since the Ranking-Hugoniot condition

$$\lambda(u_r - u_l) = f(u_r) - f(u_l)$$

must be satisfied, for a fixed value of u_l there is some restrictions on the possible values of u_r and λ. These possible values of u_r are given by the functions $S_i(\sigma)(u_l)$ which are defined in the following theorem.

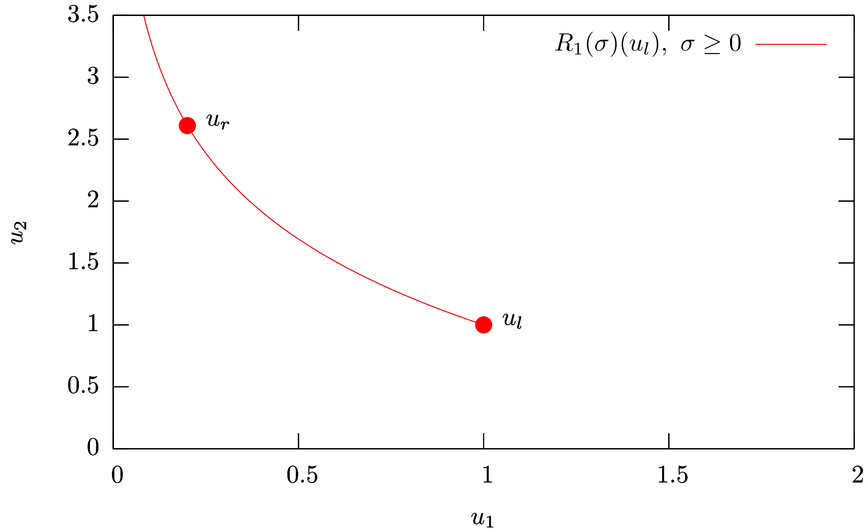
Theorem 2.4.8. *Let the system* (2.6) *be strictly hyperbolic. Then for every $u_l \in \Omega$, there exists $\sigma_0 > 0$ and n smooth curves $S_i : [-\sigma_0, \sigma_0] \mapsto \Omega$, together with n scalar functions $\lambda_i : [-\sigma_0, \sigma_0] \mapsto \mathbb{R}$ $(i = 1, \ldots, n)$, such that*

$$f(S_i(\sigma)) - f(u_l) = \lambda_i(\sigma)(S_i(\sigma) - u_l) \qquad \sigma \in [-\sigma_0, \sigma_0].$$

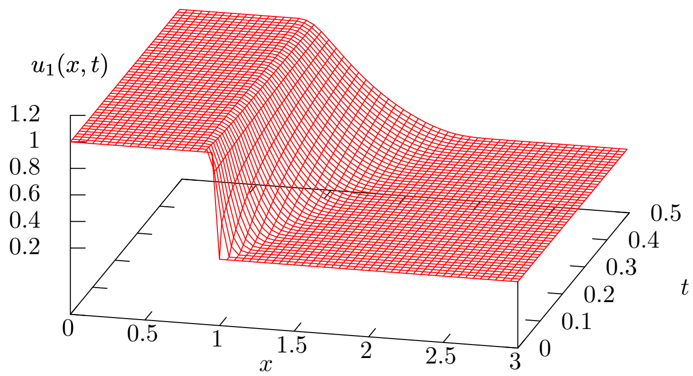
Moreover, the parametrisation can be chosen so that at $\sigma = 0$

$$\begin{aligned} S_i(0) &= u_l \\ \lambda_i(0) &= \lambda_i(u_l) \\ \left.\frac{dS_i(\sigma)}{d\sigma}\right|_{\sigma=0} &= r_i(u_l). \end{aligned}$$

This theorem expresses the fact that, for a strictly hyperbolic system, there is n curves S_i and each point on these curves can be connected to u_l by a shock wave.



(a) The point u_r is on the curve $R_1(\sigma)(u_l)$ with $\sigma \geq 0$. The solution of the Riemann problem with such initial data is a single relaxation wave.



(b) The solution of the Riemann problem. This solution is a single a relaxation wave (2.12). Except at the initial time $t = 0$, the solution is continuous.

Figure 2.4: The solution of a Riemann problem which consists of a single relaxation wave connecting to the initial states u_l and u_r.

If the i-characteristic is linearly degenerate, then the i-shock curve and the i-rarefaction curve coincide, i.e.

$$S_i(\sigma)(u_l) = R_i(\sigma)(u_l) \qquad \forall \sigma.$$

In this case, the i rarefaction wave and the i-shock wave correspond to the same solution. Indeed, consider the rarefaction solution (2.12). Since

$$r_i \bullet \lambda_i(u) = 0$$

then

$$\lambda_i(u_l) = \lambda_i(u_r)$$

and the solution (2.12) is discontinuous. This solution is sometimes called a *contact-discontinuity.*

If the i-characteristic is genuinely non-linear and the Lax admissibility condition (2.10) is imposed, then the admissible shocks are those with $\sigma < 0$. In this case, the curve $S_i(\sigma)(u_l), \sigma < 0$ can be viewed as the extension of $R_i(\sigma), \sigma > 0$ since the two curves pass by the same point u_l and have the same derivative $r_i(u_l)$ at u_l.

For example, consider the previous system of conservation laws (2.14). The Rankine-Hugoniot condition imposes that

$$\begin{pmatrix} u_r^2 u_r^1 - u_l^2 u_l^1 \\ u_r^2 y_r - u_l^2 y_l \end{pmatrix} = \lambda \begin{pmatrix} u_r^1 - u_l^1 \\ y_r - y_l \end{pmatrix}.$$

A solution is

$$\begin{aligned} u_r^2 + p(u_r^1) &= u_l^2 + p(u_l^1) \\ \lambda &= \frac{u_r^2 u_r^1 - u_l^2 u_l^1}{u_r^1 - u_l^1}. \end{aligned} \tag{2.16}$$

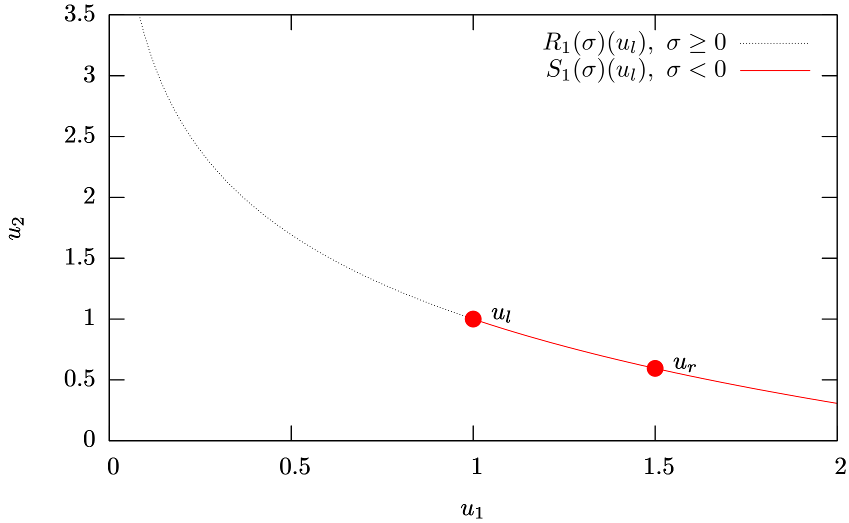
This defines the function $S_1(\sigma)$. In this particular case, $S_1(\sigma) = R_1(\sigma)$ even if the first characteristic curve is not genuinely non-linear.

If the state u_r is on the curve $S_1(\sigma)(u_l)$ (see Fig. 2.5(a)), the solution of the Riemann problem (2.6)–(2.11) is a discontinuity travelling at the speed $\lambda_i(\sigma)$ (see Fig. 2.5(b)).

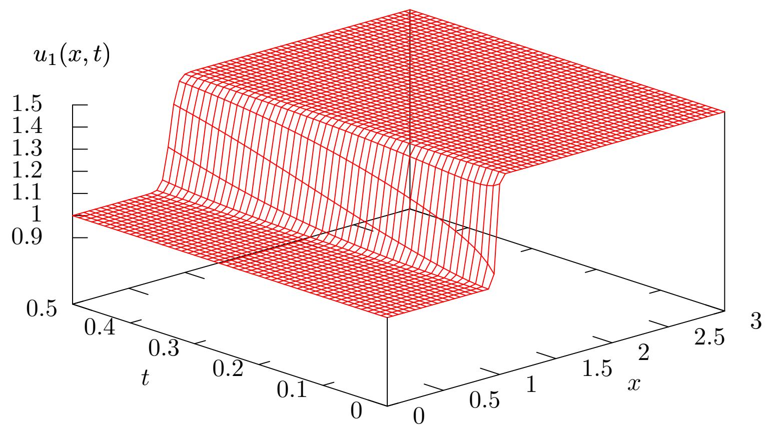
This shock wave corresponds to the situation at the tail of a traffic jam where the drivers suddenly brake when they reach it. One can show (see Section 5.1) that the system (2.13) is a traffic model where u^1 represents the speed and u^2 the density. In this case it can be noticed that the speed of the traffic jam propagation λ can easily be computed using the states after and before the discontinuity. More precisely, (2.16) indicates that λ is equal to the difference of flows divided by the difference of densities.

General solution

A general solution to the Riemann problem can be expressed in function of the two previous special cases. For given $u_l, u_r \in \Omega$, if it is possible to find



(a) The point u_r is on the curve $S_1(\sigma)(u_l)$ with $\sigma < 0$. The solution of the Riemann problem with such initial data is a single shock wave.



(b) The solution of the Riemann problem. This solution is a single shock wave. The initial discontinuity travels at a constant speed.

Figure 2.5: The solution of a Riemann problem which consists of a single shock wave connecting to the initial state u_l and u_r.

$\omega_0, \ldots, \omega_n \in \Omega$ and $\sigma_1, \ldots, \sigma_n$ such that

$$\begin{aligned}
\omega_0 &= u_l \\
\omega_i &= \begin{cases} R_i(\sigma_i)(\omega_{i-1}) & \text{if } \sigma_i \geq 0 \\ S_i(\sigma_i)(\omega_{i-1}) & \text{if } \sigma_i < 0 \end{cases} \\
\omega_n &= u_r
\end{aligned}$$

then, if the $\sigma_1, \ldots, \sigma_n$ are small enough, the global Riemann problem (2.6)–(2.11) admit a Lax-admissible weak solution which is composed by piecing together the solution of the following n basic Riemann problems

$$u(x,0) = \begin{cases} \omega_{i-1} & \text{if } x < 0 \\ \omega_i & \text{if } x > 0. \end{cases}$$

For each of these basic Riemann problems, the solution is either a rarefaction or a shock wave.

The condition on the σ_i ensures that the speed of the discontinuity (for the shock wave) or the transition range (for the rarefaction wave) remains close to λ_i. Since strictly hyperbolic systems are considered, $\lambda_1 < \ldots < \lambda_n$ and for σ_i small enough, the different waves are not overlapping.

For example, consider the already introduced system

$$\partial_t \begin{pmatrix} u^1 \\ u^2 \end{pmatrix} + \begin{pmatrix} u^2 & u^1 \\ 0 & u^2 - u^1 p'(u^1) \end{pmatrix} \partial_x \begin{pmatrix} u^1 \\ u^2 \end{pmatrix} = 0 \tag{2.17}$$

together with initial conditions

$$u(x,0) = \begin{cases} (3, 0.5)^T & \text{if } x < 0 \\ (4, 0.95)^T & \text{if } x > 0. \end{cases}$$

As explained, the solution consists in the connection of two waves. The waves associated to the first characteristic have already been explained. The description of the waves associated to the second characteristic may be obtained using the right eigenvector associated to the second eigenvalue of the matrix in (2.17):

$$\lambda_2 = u^2, \qquad r^2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

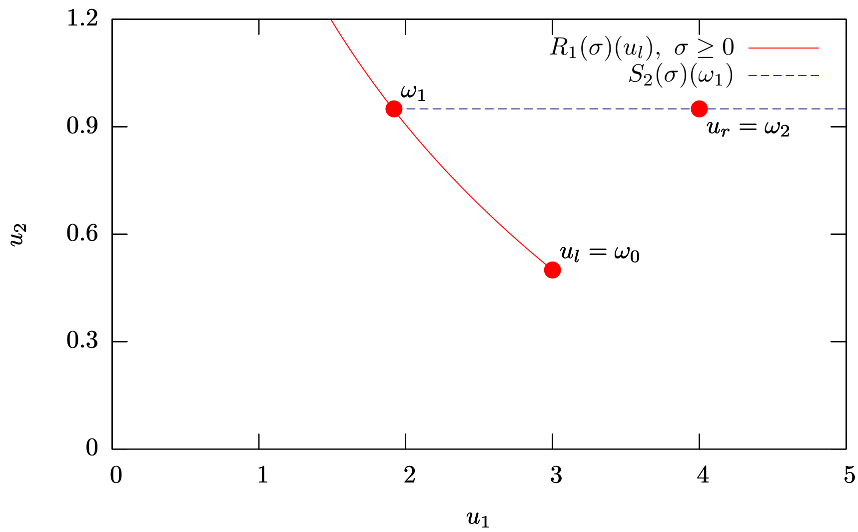
Since

$$r^2 \bullet \lambda_2 = 0 \qquad \forall u \in \Omega,$$

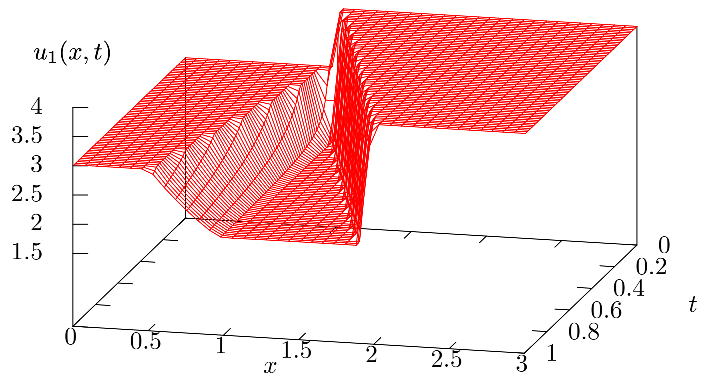
the second characteristic is linearly degenerate and its describing equation is

$$u^2 = C^{\text{st}}.$$

The given initial conditions u_l, u_r can be connected (see Fig. 2.6(a)) using a first rarefaction wave between u_l and ω_1 and then a contact discontinuity between ω_1 and u_r.



(a) The point ω_1 is on the intersection between the curve $R_1(\sigma)(u_l)$ with $\sigma \geq 0$ and $S_2(\sigma)(u_r)$. The solution of the Riemann problem with such initial data is the connection of a rarefaction between u_l and ω_1 and a contact discontinuity between ω_1 and u_r.



(b) The solution of the Riemann problem. In this solution, two basic waves are present, a rarefaction and a contact discontinuity.

Figure 2.6: The solution of a general Riemann problem which consists of the connection of a rarefaction wave (u_l, ω_1) and a contact discontinuity (ω_1, u_r).

Chapter 3

A single order network traffic model

The first model presented and extended in this thesis is the model developed by Lighthill–Whitham–Richards (LWR) [49, 60]. It is a single order model since only one PDE is used to describe the evolution of the traffic state. Two fundamental variables, function of the time t and the position on the road x, are used to describe the traffic state:

- $\rho(x,t)$: a density function representing the number of vehicles per length unit at time t and position x. For a realistic model and plausible initial condition, this variable should remain bounded between zero and a maximal density $\rho^{\max}$. This maximal density corresponds to the situation where the vehicles are bumper to bumper.
- $v(x,t)$: a speed function representing the speed of the vehicles at time t and position x. For models focused on highways, i.e. unidirectional roads, this variable should remain positive and, for reasonable drivers, be bounded by a maximal velocity $v^{\max}$.

In addition to the existence of a invariant region ($0 \leq \rho(x,t) \leq \rho^{\max}$ and $0 \leq v(x,t) \leq v^{\max}$), a realistic model should also present some properties:

- The model should be able to represent the classic observed phenomena on a highway. For example if a vehicle enters inside a traffic jam, i.e. a region where the density ρ is much higher, then the speed of the vehicle must decrease, when a vehicle leaves a traffic jam his velocity must increase, the creation and propagation of a traffic jam…
- It can also reasonably be assumed that on a road the drivers only react in function of what happens in front of them and not of what happens behind them. If the equations describing the traffic state evolution can be

written as (2.1), a condition assuring this property can easily be obtained. Recalling Remark 2.2.1, if the maximal eigenvalue of A is less than or equal to the speed v then the drivers do not react in function of what happens behind them.

The presentation of the first order traffic model is split in two parts. First the model is developed for an infinite road without in- or off-ramps. Afterwards the behaviour of the drivers at a junction connecting semi-infinite roads is modelled. This division of the network model in two parts allows to simplify the presentation. The use of an infinite road permits to introduce the main assumptions and properties of the model without having to take into account the boundary conditions. The use of a single junction with semi-infinite roads allows to analyse what happens at the junction without having to consider the whole network dynamics.

For each of these two parts the Riemann problems are analysed. One of the main reason is that if a problem with complex initial condition on the whole network is considered then the solution can be approximated through the resolution of multiple Riemann problems. A rough description of this approximation procedure is the following:

1. The initial condition is approximated by a piecewise constant approximation;
2. At each discontinuity and at the junctions the Riemann problems are solved;
3. The solution of each Riemann problem consists of waves which travel on the roads. When one of these waves hit another wave or a junction a new Riemann problem is considered at this position. Go back to step 2.

If the initial approximation is good enough then the approximated solution converges to the exact solution. See [6, 19] for a complete explanation and the restriction on the initial conditions for which the convergence is proven.

3.1 Single order model for an infinite road

The variables used to describe the traffic state are the density $\rho(x,t)$ and the speed $v(x,t)$. A traffic model is a set of equations allowing to predict the time-evolution of these variables. The vehicle conservation allows to find a first equation binding the values of ρ and v. Indeed, for a road segment $[a,b]$ without on- or off-ramp the total variation of the number of vehicles present on the segment is equal to the inflow at a minus the outflow at b

$$\frac{d}{dt}\int_a^b \rho(x,t)dx = \rho(a,t)v(a,t) - \rho(b,t)v(b,t) \tag{3.1}$$

where $\rho(x,t)v(x,t)$ is the flow at x. Assuming that $\rho(x,t)$ has continuous derivatives, one can take the limit $a \to b$ of (3.1) to obtain

$$\partial_t \rho(x,t) + \partial_x(\rho(x,t)v(x,t)) = 0. \tag{3.2}$$

This first equation alone cannot predict the time evolution of ρ and v. A second equation binding the values of ρ and v is needed. To get this second equation, the main assumption of the LWR model needs to be introduced. It is assumed that the drivers instantaneously adapt their speed in function of the surrounding density, i.e.

$$v(x,t) = V(\rho(x,t)). \tag{3.3}$$

This is of course an important simplification but not totally unrealistic if real data are observed. The data of Figure 3.1 represent the speed in function of the density on the E411 highway between Namur and Brussels. The measurements were provided by MET[1] on an hourly basis, at ten stations distributed between Daussoulx and Overijse for a period of 268 days between February 1, 2002 and October 26, 2002. A plausible $V(\rho)$ function is represented for this highway. As it can be seen, few data are available for high densities. This comes from the fact that the measurements are averaged over one hour and that the detectors do not work quite well with large densities. It is common to assume that this function $V(\rho)$ is decreasing and admits a maximal density $\rho^{\max}$ such that $V(\rho^{\max}) = 0$. This corresponds to the density where all vehicles are bumper to bumper.

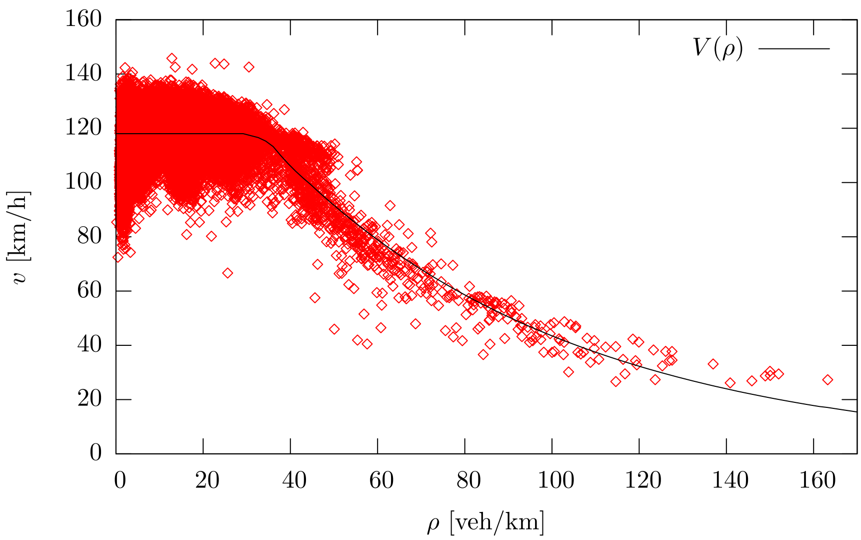


Figure 3.1: Experimental measurements of the speed as a function of the density on the E411 three-lane highway and a plausible $V(\rho)$ function.

Plugging (3.3) into (3.2), a single equation describing the time evolution is

[1] Ministry for the Equipment and Transport of the Walloon Region

obtained

$$\partial_t \rho + \partial_x f(\rho) = 0 \tag{3.4}$$

where $f(\rho) = \rho V(\rho)$ is the flow in function of the surrounding density. This function is usually assumed to be strictly concave with a single critical density σ corresponding to the maximal flow $f^{\max}$ (see Fig. 3.2 for the function f corresponding to the real data of Fig. 3.1). For densities between the critical density σ and the maximal density $\rho^{\max}$, there is congestion. It means that it is possible to have the same passing flow but with a lower density (and higher speed).

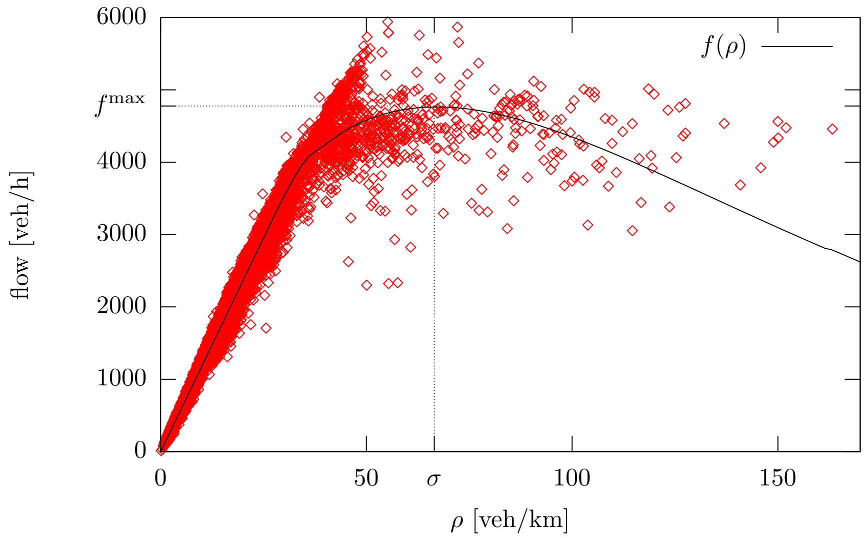


Figure 3.2: Experimental measurements of the flow as a function of the density on the E411 three-lane highway and a plausible $f(\rho)$ function.

Equation (3.4) can be rewritten as

$$\partial_t \rho + f'(\rho)\partial_x \rho = 0.$$

This equation fits in the framework of quasilinear hyperbolic systems as described in Chapter 2. Using the notations of Chapter 2 and assuming f strictly concave,

$$\begin{aligned} r_1 &= -1 \\ \lambda_1 &= f'(\rho) \\ r_1 \bullet \lambda_1 > 0. & \end{aligned}$$

This eigenvalue is thus genuinely non-linear and the solution of the Riemann problem

$$\begin{aligned} \partial_t + \partial_x f(\rho) &= 0 \\ \rho(x,0) &= \begin{cases} \rho_l & \text{if } x < 0 \\ \rho_r & \text{if } x > 0 \end{cases} \end{aligned} \tag{3.5}$$

can be expressed using a rarefaction or a shock wave.

3.1.1 Rarefaction wave

As explained in Section 2.4.4, the function $R_1(\sigma)(\rho_l),\ \sigma \geq 0$ describing all the states which can be connected to ρ_l with a rarefaction wave is the solution of the following Cauchy problem

$$\frac{du}{dt} = -1 \qquad u(0) = \rho_l.$$

Clearly, all the states ρ_r with $\rho_r \leq \rho_l$ can be connected by a rarefaction wave to ρ_l.

3.1.2 Shock wave

Two states can be connected by a shock wave only if they satisfy the Ranking-Hugoniot condition

$$f(\rho_r) - f(\rho_l) = \lambda(\rho_r - \rho_l)$$

for some λ. For $\rho_l \neq \rho_r$, the speed of the discontinuity λ is simply

$$\lambda = \frac{f(\rho_r) - f(\rho_l)}{\rho_r - \rho_l},$$

i.e. the slope of the straight line connecting the two states in the ρ–$f(\rho)$ plane (see Fig. 3.3).

The discontinuity is Lax-admissible if

$$\lambda_1(\rho_l) \geq \lambda \geq \lambda_1(\rho_r).$$

In this case, this condition becomes

$$f'(\rho_l) \geq \frac{f(\rho_r) - f(\rho_l)}{\rho_r - \rho_l} \geq f'(\rho_r).$$

Since f is assumed concave, this equation is verified for every $\rho_r \geq \rho_l$. It means that the state ρ_l can be connected by a shock wave to all the states with greater density.

The tail of a traffic jam, where the density downstream is larger than the density upstream, is the typical case where such wave is present.

3.2 Modelling of the junctions

In the previous section, a model of the evolution of the traffic state on a single road has been presented. To be able to predict the traffic evolution on a network of interconnected roads, the behaviour of the drivers at the junctions must still

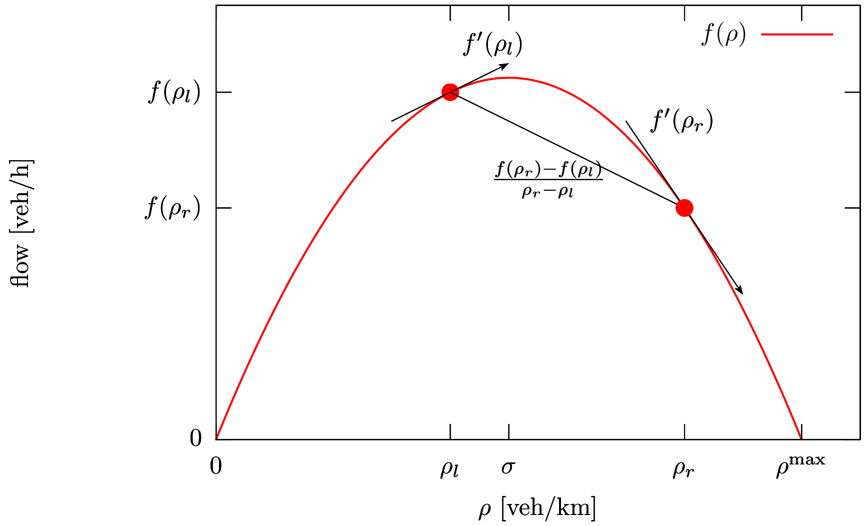


Figure 3.3: If $\rho_r \geq \rho_l$, the two states will be connected by a shock wave. The speed of the discontinuity is equal to the slope of the straight line connecting the two states.

be represented. If a road network is viewed as a graph where each arc is a road segment then the junctions correspond to the nodes. A classical junction is an intersection between different roads such as a merge or a divergence (see Fig. 3.4 (a) & (b)). Other particular junctions are the origin or end nodes (see Fig. 3.4 (c) & (d)). These nodes are used to represent the real beginning of a road or, more usually, to demarcate the part of the physical road network which is under consideration. The other junctions can be viewed as particular combinations of the previous junctions: an on-ramp is a merge junction with an origin node and a very small road, a n-input m-output crossroad is a combination of multiple merges and divergences... Since these junctions are combinations of the basic ones, only the first ones will be presented in details. Even if only the merge and diverging junctions are presented in details, various models for general n-incoming m-outgoing junction are found in the literature (see [8, 31, 35]).

The traffic evolution on an infinite road is described using a PDE and a particular problem, the Riemann problem, was analysed. Due to the intrinsic discontinuous nature of the problem at junctions, the modelling is done without introducing a new PDE. Instead it consists in the expression of the solutions for all the possible Riemann problems. Such way of modelling is useful for theoretical analysis and also for simulation since most of the numerical schemes are based on the resolution of Riemann problems. If a Riemann problem at a junction is considered, the initial state is

$$\rho_i(x,0) = \rho_{i,0} \qquad \forall x,\ \forall i \tag{3.6}$$

where subscript i refers to the semi-infinite road i. See Figure 3.5(a) for an example of a junction with two incoming and one outgoing roads.

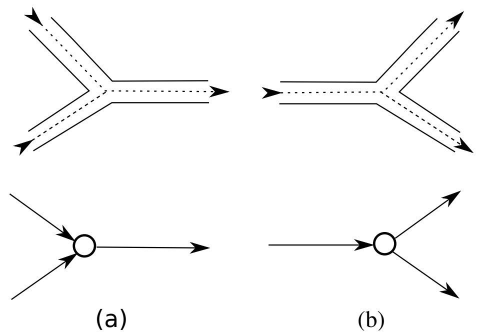


Figure 3.4: The four basic junctions: the merge (a), the divergence (b), the origin (c) and the end (d).

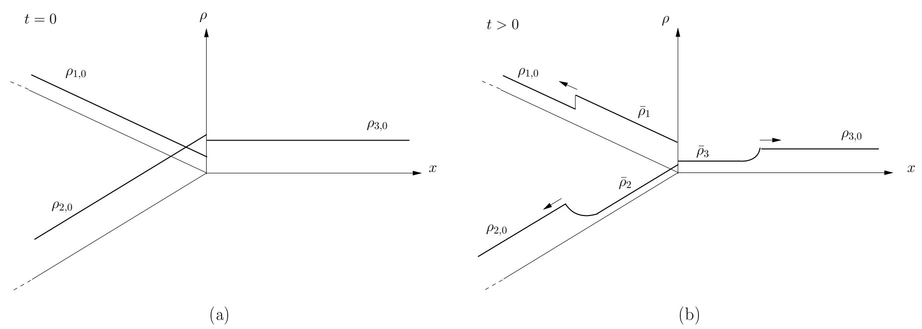


Figure 3.5: A Riemann problem for a junction with two incoming and one outgoing roads: the initial density and the density after some time. The black arrows indicate the wave directions.

Since a constant state is an equilibrium for the single road model, a modification of the state may only appear initially at the junction. Distinct Riemann problems on each road may therefore be considered:

$$\rho_i(x,0) = \begin{cases} \bar{\rho}_i & \text{if } x = 0 \\ \rho_{i,0} & \text{if } x \neq 0 \end{cases} \tag{3.7}$$

where $\bar{\rho}_i$ is the new state on the road i after the first interaction between the drivers. The set of solutions of these Riemann problems on each road (3.7) provides the solution of the junction Riemann problem (see Fig. 3.5(b)). The junction modelling problem then implies to select appropriate values $\bar{\rho}_i$ such that the collection of solutions of the Riemann problem (3.7) provides a global consistent solution to the overall Riemann problem (3.6) at the junction.

Of course, the waves produced on the incoming (resp. outgoing) roads must have a negative (resp. positive) velocity to go away from the junction in order to get a sensible model. To take this constraint into account in the model, the set of possible values of $\bar{\rho}_i$ is restricted to a subset of $\mathbb{R}$ called the *admissible region* such that all waves produced by the Riemann problems (3.7) have negative (resp. positive) speed if i corresponds to an incoming (resp. outgoing) road.

A frequently asked question is: "why is the density at the junction constant?". The answer is in two parts:

1. A *single* junction with *semi-infinite* roads is considered. If a wave emitted from the junction is travelling on a road, the speed of this wave is constant (see Section 2.4.4) and will thus never hit back the junction. If a network was considered the situation would have been different. In that case a wave emitted from a junction can hit another junction and re-modify the density at this second junction.
2. Mathematically a constant density near the junction is the only solution such that, on each road, there is a possible admissible weak solution. This fact has already been presented in Section 2.4.4. Indeed, for the special solutions of the Riemann problem (the rarefaction and the shock wave) the state $u(x,t)$ is constant at $x = 0$ which corresponds to the position of the initial discontinuity.

This section is structured as follow: first the admissible regions for incoming and outgoing roads are described and then additional conditions are presented in order to have a unique physically acceptable solution to the problem of selecting $\bar{\rho}_i$.

3.2.1 The admissible regions for the junction model

As already mentioned, in order to formulate a junction model it is needed to explicit, for each road, the admissible regions for the values of $\bar{\rho}_i$.

Incoming road

For an incoming road, the wave present in the solution of the Riemann problem has a negative speed. Due to the speed of the wave described in Section 3.1, it can be shown that the new state $\bar{\rho}_i$ belongs to the following particular set (see Fig. 3.6).

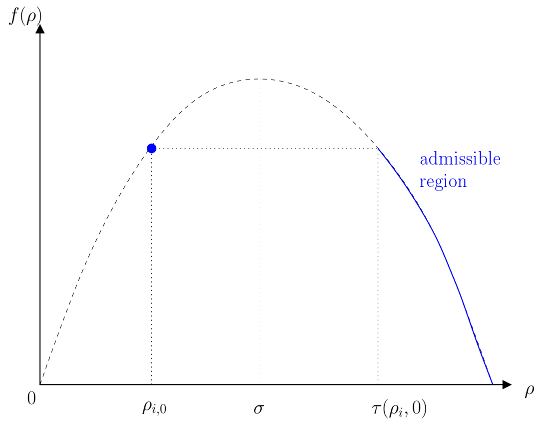
$$\bar{\rho}_i \in \begin{cases} \{\rho_{i,0}\} \cup]\tau(\rho_{i,0}), \rho_{\max}] & \text{if } \ 0 \le \rho_{i,0} < \sigma \\ [\sigma, \rho_{\max}] & \text{if } \ \sigma \le \rho_{i,0} \le \rho_{\max} \end{cases}$$

where, for each $\rho \neq \sigma$, $\tau(\rho)$ is the unique number $\tau(\rho) \neq \rho$ such that

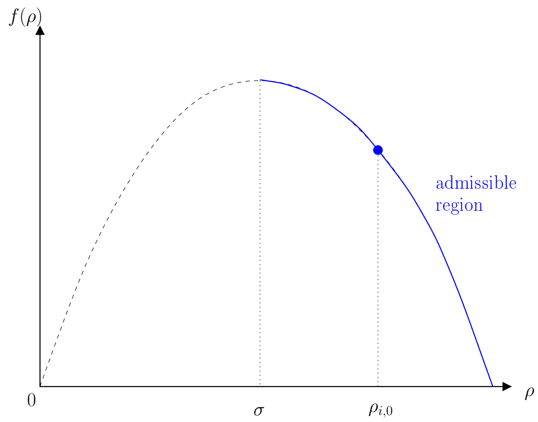
$$f(\rho) = f(\tau(\rho)).$$

In the first case where $\rho_{i,0} \le \sigma$, the "left state" $\rho_{i,0}$ will be connected by a shock wave with negative speed to $\bar{\rho}_i$ (see Fig. 3.6(a)).

In the second case where $\rho_{i,0} \ge \sigma$, two sub-cases must be distinguished:



(a) $\rho_{i,0} \leq \sigma$



(b) $\rho_{i,0} > \sigma$

Figure 3.6: The admissible regions for an incoming road.

- if $\bar{\rho}_i \geq \rho_{i,0}$ the solution consists of a shock wave with a negative speed;
- if $\bar{\rho}_i \leq \rho_{i,0}$ the solution will be a rarefaction wave. In order that the right limit of this rarefaction wave has a negative speed, we need that $\bar{\rho}_i \geq \sigma$ (see Fig. 3.6(b)).

This restriction on the possible values of $\bar{\rho}_i$ can be expressed in function of the flow:

$$f(\bar{\rho}_i) \in [0, S_i] = \begin{cases} [0, f(\rho_{i,0})] & \text{if } 0 \leq \rho_{i,0} \leq \sigma \\ [0, f(\sigma)] & \text{if } \sigma \leq \rho_{i,0} \leq \rho_{\max} \end{cases} \tag{3.8}$$

where S_i represents the "sending" capacity (sometimes called the traffic demand) of an incoming road, i.e. the maximal flow which can leaves the incoming road (see Fig. 3.7). This notion has first been introduced in the numerical scheme of the classical Riemann problem (3.5) (see [12]) and then used for the description of the junctions (see [11] and [44]).

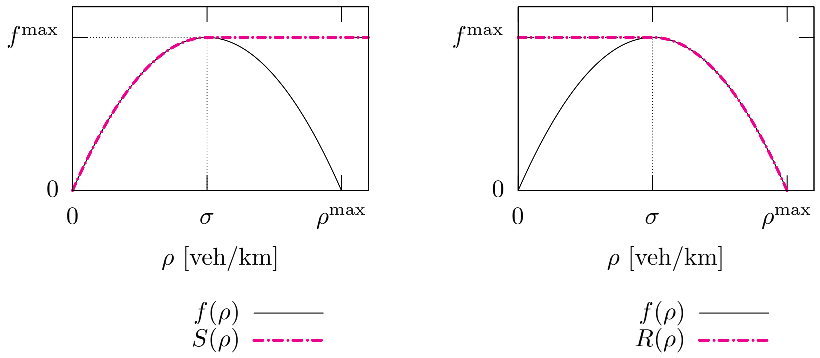


Figure 3.7: The sending (left) and receiving (right) functions used in the description of the admissible regions.

Outgoing road

For an outgoing road, the wave present in the solution of the Riemann problem has a positive speed. It can be shown that the new state $\bar{\rho}_i$ belongs to the following particular set

$$\bar{\rho}_i \in \begin{cases} [0,\sigma] & \text{if} \quad 0 \le \rho_{i,0} \le \sigma \\ \{\rho_{i,0}\} \cup [0,\tau(\rho_{i,0})[& \text{if } \sigma < \rho_{i,0} \le \rho_{\max}. \end{cases}$$

This set can be viewed as the symmetric of the incoming road admissible region. The graph describing this region can be obtained from the Figure 3.6 applying an axial symmetry around $\rho = \sigma$.

In the first case where $\rho_{i,0} \le \sigma$, two sub-cases must be distinguished:

- if $\bar{\rho}_i \le \rho_{i,0}$ the solution consists of a shock wave with a positive speed;
- if $\rho_{i,0} \le \bar{\rho}_i$ the solution will be a rarefaction wave. In order that the left limit of this rarefaction wave has a positive speed, it is needed that $\bar{\rho}_i \le \sigma$.

In the second case where $\rho_{i,0} \ge \sigma$, the "right state" $\rho_{i,0}$ will be connected by a shock wave with positive speed to $\bar{\rho}_i$.

This restriction on the possible values of $\bar{\rho}_i$ can be expressed in function of the flow.

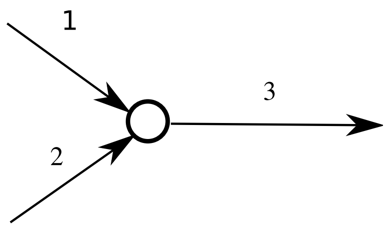
$$f(\bar{\rho}_i) \in [0, R_i] = \begin{cases} [0, f(\sigma)] & \text{if } 0 \le \rho_{i,0} \le \sigma \\ [0, f(\rho_{i,0})] & \text{if } \sigma \le \rho_{i,0} \le \rho_{\max} \end{cases} \tag{3.9}$$

where R_i represents the "receiving" capacity (sometimes called the traffic supply) of an outgoing road, i.e. the maximal flow which can enters the outgoing road (see Fig. 3.7).

3.2.2 Additional conditions

After the presentation of the admissible regions limiting the possible values of $\bar{\rho}_i$, it remains to express what are the appropriate values. Since the merge junction is the most complex one, it will be completely described while the others (divergence, origin and end) will be more quickly reviewed.

The merge junction



A first restriction on the values of the densities at the borders of the junction is the conservation of flows:

$$\sum_{i=1}^{2} f(\bar{\rho}_i) = f(\bar{\rho}_3).$$

This restriction is not sufficient since an infinity of values inside the admissible region satisfy this conservation of flows. A natural additional condition is a maximisation function of the passing flow (see [35, 8, 43, 47, 38, 11]). There are mainly two approaches to this maximisation. The first one is simply the maximisation of the total passing flow:

$$\max_{\bar{\rho}_i} \sum_{i=1}^{2} f(\bar{\rho}_i).$$

The second one is more complex:

$$\max_{\bar{\rho}_i} \sum_{i=1}^{2} \Phi_i(f(\bar{\rho}_i))$$

where Φ_i is increasing and strictly concave. The main ideas behind these two approaches are:

1. Each driver tries to pass the junction if it is possible. If the flow was not maximised, this assumption would not be satisfied.
2. From a mathematical point of view, selecting the solution corresponding to a maximisation of the flows allows to have the same solution for a Riemann problem on an infinite road and for a Riemann problem on an artificial 1-incoming–1-outgoing road.

First approach: $\max_{\bar{\rho}_i} \sum_{i=1}^{2} f(\bar{\rho}_i)$**.** This first approach appears simpler but has some drawbacks when the "priority factors" must be introduced (see below).

Taking into account the restrictions on the values $\bar{\rho}_i$, the model for the merge junction can be expressed in terms of flow as

$$\begin{array}{rcl} \max\limits_{\bar{\rho}_i} & & \sum\limits_{i=1}^{2} f(\bar{\rho}_i) \\ f(\bar{\rho}_1) + f(\bar{\rho}_2) & = & f(\bar{\rho}_3) \\ f(\bar{\rho}_1) & \in & [0, S_1] \\ f(\bar{\rho}_2) & \in & [0, S_2] \\ f(\bar{\rho}_3) & \in & [0, R_3]. \end{array}$$

In the case where $S_1 + S_2 \leq R_3$, the maximum is unique

$$\begin{cases} f(\bar{\rho}_i) = S_i & i = 1, 2 \\ f(\bar{\rho}_3) = S_1 + S_2. \end{cases}$$

If $S_1 + S_2 > R_3$, there are too many vehicles trying to enter the junction for the available space on the outgoing road. In this case, some "priority factors" between the incoming flows must be given. These factors, which are not easily formulated, express how the available space is shared between the incoming drivers. Several factors have been proposed in the literature (see [38]):

1. The priority factors can be fixed (such as in [8])

$$\alpha_i = C^{st}.$$

2. The priority factors may depend on some fixed coefficients p_i depending on the road geometry (such as in [38]):

$$\alpha_i = \begin{cases} p_i & R_3 p_i \leq S_i \\ \frac{S_i}{R_3} & S_i < R_3 p_i. \end{cases}$$

3. Another intuitively priority factor used is the proportion of the incoming flow wishing to enter the outgoing road:

$$\alpha_i = \frac{S_i}{\sum_{i=1}^{2} S_i}.$$

The possible passing flow is then split between the incoming roads in function of these priority factors:

$$\begin{cases} f(\bar{\rho}_i) = \alpha_i R_3 & i = 1, 2 \\ f(\bar{\rho}_3) = R_3. \end{cases}$$

Unfortunately each of these priority factors have some important drawbacks. In the first two cases, there exist situations where a traffic jam occurs on one of the incoming roads while the outgoing road remains uncongested. In the third case, even if the priority factors are intuitively acceptable, there exist initial conditions such that there is no solution to the Riemann problem. See [47] for such an example and a condition on the priority factors to ensure the existence of a solution.

Second approach: $\max_{\bar{\rho}_i} \sum_{i=1}^{2} \Phi_i(f(\bar{\rho}_i))$**.** Using this second approach, the problems of the previous approach can be avoided. The model can be rewritten as

$$\begin{array}{rcl} \max\limits_{\bar{\rho}_i} & & \displaystyle\sum_{i=1}^{2} \Phi_i(f(\bar{\rho}_i)) \\ f(\bar{\rho}_1) + f(\bar{\rho}_2) & = & f(\bar{\rho}_3) \\ f(\bar{\rho}_1) & \in & [0, S_1] \\ f(\bar{\rho}_2) & \in & [0, S_2] \\ f(\bar{\rho}_3) & \in & [0, R_3]. \end{array}$$

It can be shown (see [35]) that the solution always exists and is unique. The "priority factors" which were difficult to express in the previous models are already taken into account by the Φ_i-functions.

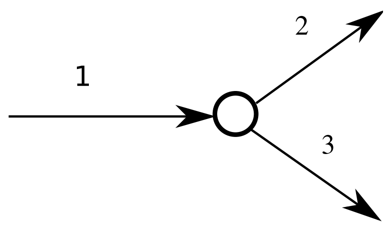
It can also be shown (see [47]) that there exists a correspondence between this approach and another approach where an additional variable is introduced to describe what happens inside the junction.

A particular function Φ_i which can be used is

$$\Phi_i(f(\bar{\rho}_i)) = f(\bar{\rho}_i) - \frac{f(\bar{\rho}_i)^2}{2 p_i f^{\max}}.$$

This function leads to the merge model proposed in [11] which has a nice interpretation: "*As long as the supply of vehicles from both roads, S_1 and S_2, is not exhausted, it is assumed that a fraction (p_1) of the vehicles comes from road 1 and the remainder (p_2) from road 2, where $p_1 + p_2 = 1$. The constants p_i are characteristics of the intersection and reflect any priorities. (Absolute priority from road 1, for example, would imply that $p_1 = 1$ and $p_2 = 0$.) If the supply of vehicles on one of the incoming roads is exhausted then the remaining vehicles comes from the other road.*"

The diverging junction



It is natural to assume that, the final destinations of the drivers being fixed, there exists a proportion α of the vehicles from road 1 wishing to enter road 2. With this assumption, the flow leaving the first road to enter the second is

$$F_2 = \min(\alpha S_1, R_2)$$

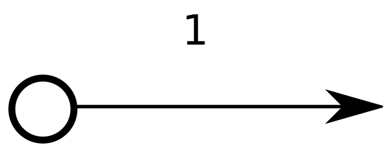
and the flow entering the third road is

$$F_3 = \min((1-\alpha)S_1, R_3).$$

After having obtained the flows, the new densities at the borders of the junction are chosen such that

$$\begin{aligned} f(\bar{\rho}_1) &= F_2 + F_3 \\ f(\bar{\rho}_2) &= F_2 \\ f(\bar{\rho}_3) &= F_3. \end{aligned}$$

The origin junction



The origin junction is used to inject vehicles into the network. If the flow wishing to enter the network is described by the function $D(t)$, a first commonly used approach consists in expressing the partial derivative equation together with a boundary condition involving $D(t)$:

$$\begin{aligned} \partial_t \rho + \partial_x f(\rho) &= 0 \qquad x > 0 \\ \rho(0,t) &= \rho_0(t) \end{aligned}$$

with

$$f(\rho_0(t)) = D(t).$$

Unfortunately this approach does not ensure the existence and unicity of the solution for all functions $D(t)$ (see [3, 43]). To remove this important limitation, we propose a new model for the origin junction.

A virtual road segment of length Δx used as a buffer is introduced. This buffer stores the flow wishing to enter the road but which has not yet been able to do it. This can be the case if road 1 is too much congested. Denoting by ρ_b the density in the buffer and R the receiving capacity of the outgoing road, the flow entering the road is

$$F = \min(S(\rho_b), R)$$

and the dynamics of the density of the buffer is

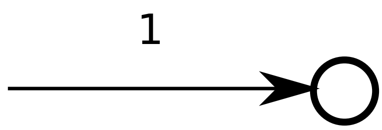
$$\frac{d\rho_b}{dt} = \frac{1}{\Delta x}(D(t) - F).$$

The new state ρ_1 on the outgoing road is the only density such that

$$\begin{aligned} f(\bar{\rho}_1) &= F \\ \bar{\rho}_1 &\in \text{admissible region of road 1.} \end{aligned}$$

The introduction of the new state variable ρ_b allows to remove the restriction on $D(t)$.

The end junction



The end junction is used to represent a final destination of some vehicles. When vehicles leave the last road before the junction, they definitively disappear from the system. For the modelling of the end junction, it may be assumed that the road is not influenced by what is after the junction, i.e. imposing a Neumann condition ($\frac{\partial \rho}{\partial x} = 0$) at the end of the road. Imposing a Neumann condition correspond to have

$$F = f(\rho_{1,0})$$

where F is the flow leaving the last cell. In some case, it may be useful to impose an upper bound on the flow $F_{\max}$. For example, it may be known that in the "real world" the vehicles leaving the system via the end junction enter an already congested road which may only accept a maximum flow of $F_{\max}$. In this case, the flow leaving the last road is

$$F = \min(f(\rho_{1,0}), F_{\max}).$$

This flow allows to select the new value $\bar{\rho}_1$.

3.2.3 The capacity drop phenomenon

The *capacity drop phenomenon* is a critical phenomenon which represents the fact that the outflow of a traffic jam is significantly lower than the maximum achievable flow at the same location. We can easily understand this phenomenon in the case of a merge junction where two roads merge in one: if there are too many vehicles trying to access the same road, there is a sort of mutual hindrance between the drivers which results in an outgoing flow lower than the optimal possible flow.

This phenomenon has been experimentally observed (see [7] and [22]). The flow decrease, which may range up to 15 %, has a considerable influence when considering traffic control ([52]). To have of a model describing this phenomenon is thus a critical element in the establishment of a traffic state regulation strategy.

Unfortunately the criterion described in the previous paragraph which consists of maximising the passing flow:

$$\sum_{i=1}^{2} f(\bar{\rho}_i)$$

under the constraints (3.8)–(3.9) does not represent the capacity drop phenomenon. Instead of maximising the criterion over the region defined by (3.8)–(3.9), it is possible to do the maximisation over a subregion. Defining

$$R'_3 = \min\left(R_3, g\left(\sum_i S_i\right)\right) \tag{3.10}$$

where $g(x)$ is a function whose shape is represented in Figure 3.8. In this expression of R'_3, we have

- $\sum_i S_i$ which represents the sum of the flows wishing to enter the outgoing road;
- the function $g(\cdot)$ which expresses the fact that when too many vehicles are trying to enter the same road ($\sum_i S_i > f(\sigma)$), there is a sort of mutual obstruction which decreases the capacity of the outgoing road ($R'_3 \leq g(\cdot) < f(\sigma)$);
- the $\min(R_3, \cdot)$ to be guaranteed to remain in a subregion of (3.9).

The total passing flow may now be maximised with any priority factors described in the previous paragraph with

$$f(\bar{\rho}_i) \in \begin{cases} [0, S_i] & i = 1, 2 \\ [0, R'_3] & i = 3 \end{cases}$$

instead of (3.8)–(3.9).

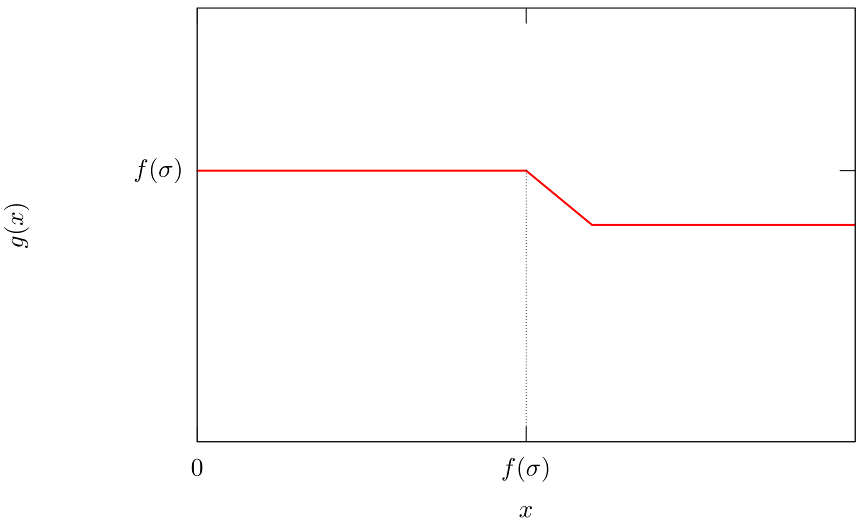
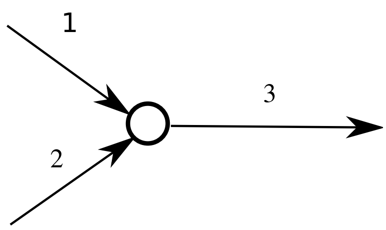


Figure 3.8: Possible shape of a g-“capacity drop function”.

This redefinition of the reception capacity is an appropriate expression of the *capacity drop phenomenon*: if too many drivers are trying to access the same road (a traffic jam is occurring), the output flow will be lower than the maximum achievable flow ($f(\sigma)$). This phenomenon has an important influence on the global behaviour of the system and should thus be taken into account when performing simulations. Another approach to represent this phenomenon is to introduce an additional variable which represents the state of the junction (see [44]).

This strong influence is illustrated in the following simple simulation. The network simply consists of two roads merging into one.



The function $f(\rho)$ used has a maximal value of $f^{\max} = 2880$. As initial condition, we take 1400 [veh/h] for the flows on the two incoming roads and 2800 on the outgoing one. These are admissible states for (3.8)–(3.9) and it is optimal for the additional criterion presented in Section 3.2.2. After a while, a temporary increase of flow up to 1550 is added at the beginning of the first incoming road (see Fig. 3.9). The sum of the two incoming flows (1400+1550=2950) is too high for the outgoing road so a traffic jam will occur. Two cases may be distinguished:

without the capacity drop representation : the traffic jam will grow and, when the incoming flow on the first road will finally go back to 1400, the traffic jam will decrease to finally disappear (see Fig. 3.10 where the traffic state on one of the incoming roads and on the outgoing one are represented).

with the capacity drop representation : because of the presence of the traffic jams in front of the junction, the reception capacity of the outgoing road will drop down to 2736 ($= 0.95 f(\rho^*)$). The traffic jam will grow and, even after that the incoming flow on the first road has gone back to 1400, the sum of the incoming flows is too high for the new reception capacity of the outgoing road ($1400 + 1400 > 2736$). As it can be seen in Fig. 3.11, the traffic jam on the incoming roads will never stop of growing.

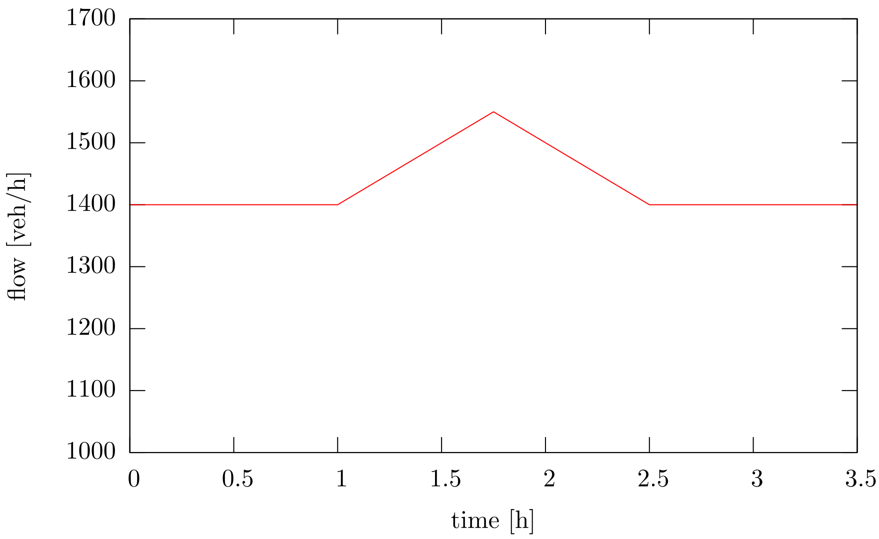


Figure 3.9: The flow at the entrance of the first incoming road.

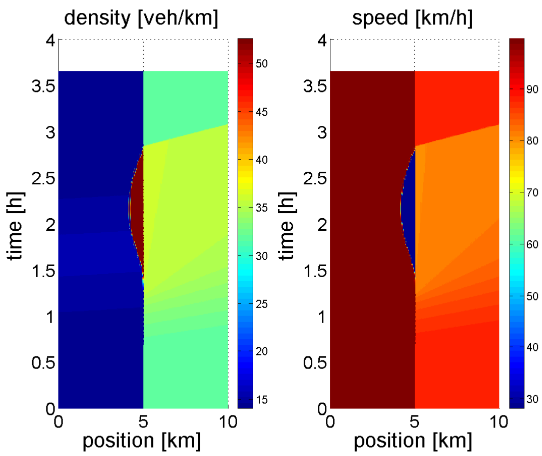


Figure 3.10: The evolution of the traffic state in absence of a capacity drop representation. The first five kilometres correspond to the first incoming road and the second five kilometres to the outgoing road.

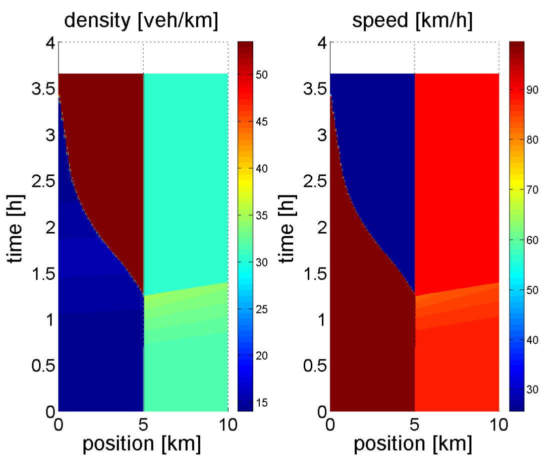


Figure 3.11: The evolution of the traffic state in presence of a capacity drop representation. The first five kilometres correspond to the first incoming road and the second five kilometres to the outgoing road.

3.3 Numerical simulator

A numerical simulator has been developed in order to validate the LWR model and various control strategies. For such simulator, it may be useful to make a distinction between different *classes* of vehicles. For example the vehicles with the same origin–destination path belong to the same class. Being able to distinguish the vehicles in function of their path is very useful to describe the behaviour of the traffic at the junctions between different roads. If there are n different classes the LWR model (3.4) becomes

$$\frac{\partial}{\partial t}\begin{pmatrix}\rho^1\\ \vdots\\ \rho^n\end{pmatrix} + \frac{\partial}{\partial x}\begin{pmatrix}\frac{\rho^1}{\rho^{\text{tot}}} f(\rho^{\text{tot}})\\ \vdots\\ \frac{\rho^n}{\rho^{\text{tot}}} f(\rho^{\text{tot}})\end{pmatrix} = 0 \tag{3.11}$$

where $\rho^i(x,t)$ is the density of vehicle of class i in position x at time t and $\rho^{\text{tot}} = \sum_i \rho^i$ the total density.

3.3.1 General overview

The simulator was developed in C++ for an arbitrary road network based on the model (3.11) and the junction modelling described in Section 3.2. The distinction between classes is based upon origin–destination paths. Only one main $f(\rho)$ function may be specified by the user for the whole network. By convention, this $f(\rho)$ function represents the flow–density relation for a three-lane road without additional speed limitation. If the user specifies a different number of lanes and/or a speed limitation for an arc, the flow–density relation used for this arc is constructed from the main $f(\rho)$ relation by the following transformation:

$$f_i(\rho) = \rho \min(v_{\max}, \underbrace{\frac{i}{3}\frac{f\left(\frac{3}{i}\rho\right)}{\rho}}_{U_i(\rho)})$$

where i is the number of lanes and $v_{\max}$ the maximum allowed speed.

The structure of the program is sufficiently flexible to easily implement various control strategies such as ALINEA [51], user-optimal routing [59], system-optimal routing (see Section 4.3). It means that some properties (maximum allowed speed, the maximum flows leaving the origin junction,...) may be updated during the simulation.

3.3.2 Numerical scheme

In order to solve the partial differential equation (3.4), an appropriate numerical scheme is needed. Different schemes exist but the most simple and intuitive is the Daganzo scheme (see [12, 42]).

In this scheme, the highway is divided into cells of equal size Δx. The scalar $\rho_i(k)$ represents the average density in cell i at time $k \cdot \Delta t$ where Δt is the discrete time step. The function $f(\rho)$ is assumed to have only one maximum $f^{\max}$ at the critical density σ. The idea of this numerical scheme is to compute the flows between the cells by assuming that the densities in each cells are constant (equal to the average densities).

The flow between the cell i and the cell $i+1$ can be computed by solving the following Riemann problem

$$\rho(x,0) = \begin{cases} \rho_i & \text{if} \quad x \leq 0 \\ \rho_{i+1} & \text{if} \quad x > 0 \end{cases} \tag{3.12}$$

and taking the flow at $x = 0, t > 0$. The value of this flow is equal to

$$F_{i+\frac{1}{2}} = \min(S(\rho_i), R(\rho_{i+1})). \tag{3.13}$$

where $S(\rho)$ and $R(\rho)$ are the functions defined in Section 3.2.1 as

$$\begin{aligned} S(\rho) &= \begin{cases} f(\rho) & \text{if } \rho \leq \sigma \\ f^{\max} & \text{if } \rho > \sigma \end{cases} \\ R(\rho) &= \begin{cases} f^{\max} & \text{if } \rho \leq \sigma \\ f(\rho) & \text{if } \rho > \sigma \end{cases}. \end{aligned}$$

The expression (3.13) can be intuitively understood by considering the Riemann problem (3.12) as a junction problem with one incoming and one outgoing road. This Riemann junction problem is then expressed as

$$\begin{aligned} \max \quad & f_1 \\ f_1 &= f_2 \\ f_1 &\in [0, S(\rho_i)] \\ f_2 &\in [0, R(\rho_{i+1})] \end{aligned}$$

and the solution is clearly given by (3.13).

After having computed the flows between the cells, the densities are updated according to the rule:

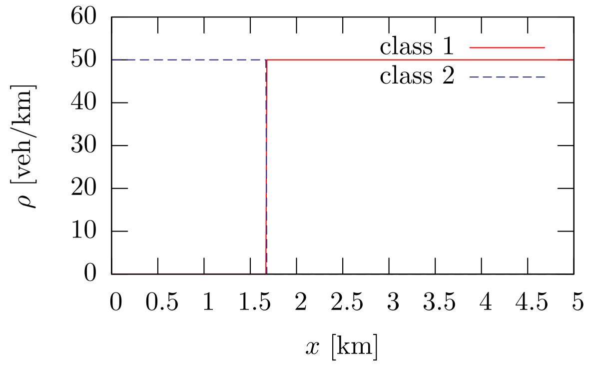
$$\rho_i(k+1) = \rho_i(k) + \frac{\Delta t}{\Delta x}\left(F_{i-1/2} - F_{i+1/2}\right).$$

If the LWR model with different classes is considered, the Daganzo scheme can easily be extended in order to solve (3.11). First, the flows between the different cells are computed by (3.13) where the sending and receiving capacities are evaluated on the basis of the total density $\rho^{\text{tot}} = \sum_j \rho^j$. Then, the density ρ_i^j is updated as follows:

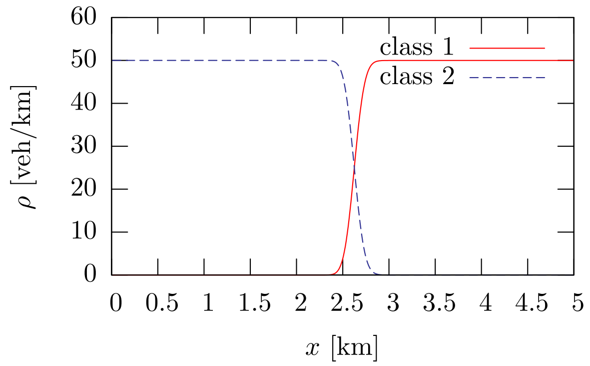
$$\rho_i^j(k+1) = \rho_i^j(k) + \frac{\Delta t}{\Delta x}\left(F_{i-1/2}\frac{\rho_{i-1}^j(k)}{\sum_j \rho_{i-1}^j(k)} - F_{i+1/2}\frac{\rho_i^j(k)}{\sum_j \rho_i^j(k)}\right).$$

This extended Daganzo scheme is easy to implement but doesn't converge to the exact solution of (3.11) even when $\Delta x \to 0$. One may consider the particular initial conditions represented in Figure 3.12(a) where there are only two classes of vehicles. A single road is considered where on the first part there are only vehicles of class 2. On the second part there are initially only vehicles of class 1. The initial density being constant everywhere, all the vehicles travel at the same speed, none can overtake another and the exact solution must just be the displacement of the discontinuity. The solution from the extended Daganzo scheme is represented in Figure 3.12(b). We can observe the presence of some numerical diffusion in this extended Daganzo scheme. A more complex scheme can be used which does not exhibit such numerical diffusion. For example the results of a simulation using the WENO (Weighted Essentially Non-oscillatory) procedure (see [57]) for the reconstruction of the density at the boundaries of the cells and the Lax-Friedrichs flux for the approximation of flow at the boundaries are represented in Figure 3.12(c). It can be noticed that the numerical diffusion is less important.

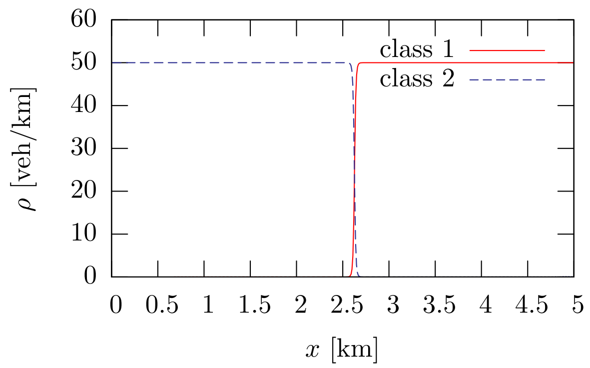
In this numerical simulator the extended Daganzo scheme is used even under the presence of the numerical diffusion. This choice is based on two reasons. First, the Daganzo scheme is much simpler and allows faster computation. Second, the presence of the diffusion is not a big problem. Indeed the evolution of the total density is correctly evaluated and in reality, all the drivers don't have the same behaviour. It can thus be argued that the presence of some numerical diffusion adds a little realism.



(a) The initial condition of the Riemann problem: there are only vehicles of class 2 on the first part of the road and only vehicles of class 1 on the second part.



(b) The state predicted by the extended Daganzo scheme a few time after the start of the simulation. The presence of numerical diffusion can be noticed.



(c) The state predicted by the WENO scheme a few time after the start of the simulation. There is less numerical diffusion than using the extended Daganzo scheme.

Figure 3.12: Two numerical simulations on a single road. There are two classes of vehicles which are initially not mixed.

3.3.3 The user inputs

The user may encode the topology of the network used in the simulator using a graph editor (see Fig. 3.13). He can also specify, using this graph editor, some parameters such as:

- the length of the arcs;
- the number of lanes of the arcs;
- the maximal allowed speed on the arcs;
- the type of the junction (in the case of a node with one incoming and one outgoing arc, there exist different possibilities);
- the maximal allowed output flow rate in case of an end or off-ramp junction.

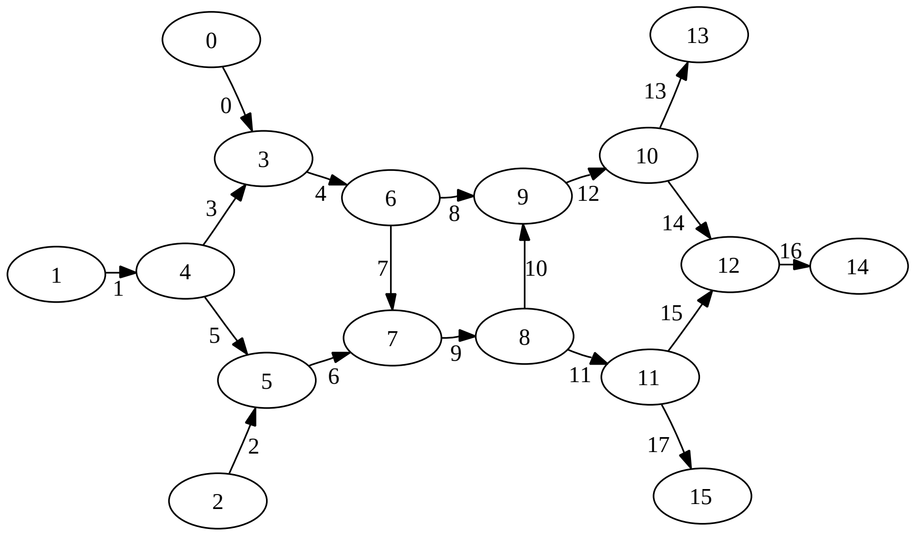


Figure 3.13: The encoding of the network topology.

The other parameters are specified using text files:

1. The different classes of vehicles considered by specifying the origin–destination paths. If the user encodes only the origin and the destination, the shortest path (in length) will be assumed.
2. The flow demand at the entrances (origins and on-ramps) which may be time varying.
3. The remaining parameters:

- the length of the simulation;
- the time step Δt used in the numerical scheme;
- the length step Δx used in the numerical scheme;
- if the capacity drop phenomenon is active;
- if some noise with zero mean is added to the flow described by the user.

3.3.4 The simulator outputs

The evolution of the density–flow–speed on each arc are plotted and displayed (see Fig. 3.14) during the simulation. Moreover the data are written in text files for possible post-processing. At the same time, various costs are computed

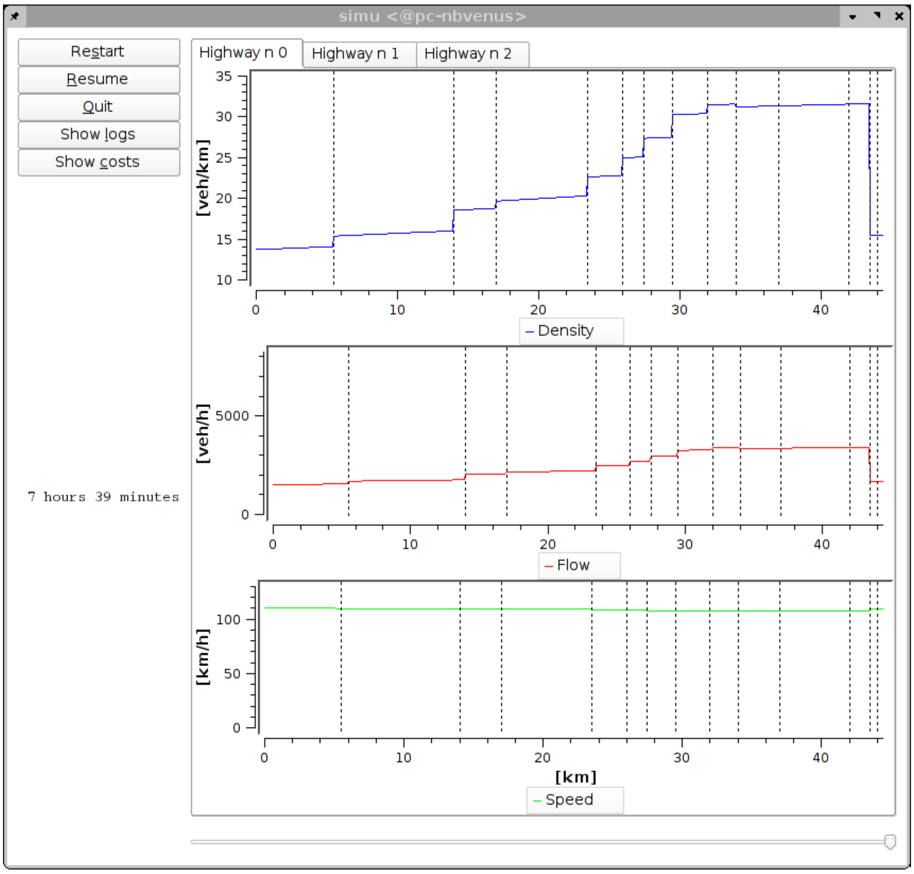


Figure 3.14: The evolution of the density–flow–speed during the simulation.

in order to evaluate different control strategies. These costs are derived in Section 4.1. The values of the total costs are updated and displayed during the simulation. Moreover the cost rates are graphically displayed allowing the user to see the effect on the total costs of the different events (see Fig. 3.15). Different costs are taken into account: a time cost involving nominal cost and overtime cost, the Total Time Spent and the pollution cost.

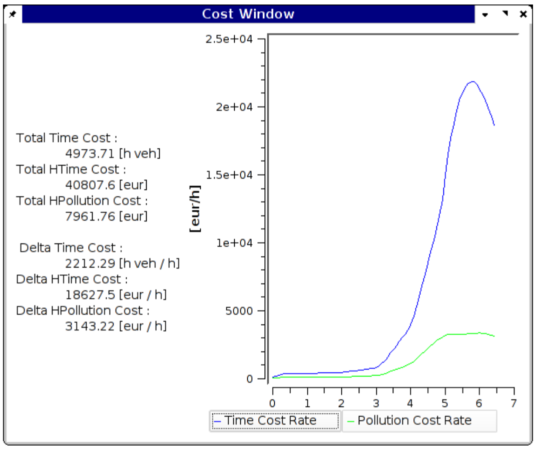


Figure 3.15: The evolution of the costs during the simulation.

3.3.5 Simulation of the E411 highway

Since traffic measurements are available for the E411 highway, the numerical simulator can be used to validate the LWR model. The measurements were provided by MET on a minute basis, at ten stations distributed between Daussoulx and Overijse for a period of 31 days between October 1, 2003 and October 31, 2003. See Fig. 3.16 and Tab. 3.1 for a scheme of the network and the position of the traffic sensors. The minutely basis allows to obtain more information for high densities than with the hourly data presented in Figure 3.1. It also allows to obtain precise information on the in/out-flow at the ramps. The mean speed in function of the density and the 90% interval confidence (assuming the speed normally distributed) are represented in Figure 3.17. In the same figure a plausible $V(\rho)$ relation is represented

$$V(\rho) = \begin{cases} -34.4 \operatorname{atan}(0.036(\rho - 96.6)) + 67.2 & \text{if } \rho \leq 150 \\ \frac{5976.7}{\rho} - 10 & \text{if } \rho > 150. \end{cases}$$

For low densities the coefficients were obtained using a mean-square minimisation. For high densities, where there is less information, an extrapolation was performed leading to a plausible maximal density $\rho^{\max} = 600$ [veh/km] for a three-lane highway. This function $V(\rho)$ leads to a plausible maximal flow $f^{\max} = 6867$ [veh/h].

Location of the station	Station number	Location
Daussoulx - Eghezée [12]	25	1
Daussoulx - Eghezée [12]	23	4.85
Eghezée [12] - Thorembais-St-Trond [11]	21	6.74
Eghezée [12] - Thorembais-St-Trond [11]	19	13.13
Thorembais-St-Trond [11] - Walhain [10]	17	15.15
Walhain [10] - Corroy-le-Grand [9]	15	22.6
Corroy-le-Grand [9] - Louva-la-Neuve [8a]	13	24.45
Louvain-la-Neuve [8a] - Louvranges [6]	11	26.8
Wavre [6] - Bierges [5]	9	30.65
Bierges [5] - Rosières [4]	7	33.15
Rosières [4] - Overijse [3]	5	35.15

Table 3.1: The position of the counting stations on the E411 highway

It can be noticed that the function $f(\rho)$ represented in Figure 3.18 is not strictly concave. It means that this function does not fit perfectly in the framework developed in Chapter 2; more precisely the eigenvalue $f'(\rho)$ is not genuinely non-linear. Some results presented in Section 2.4.4 are thus not applicable in this case. For example, if the function $f(\rho)$ is strictly concave then it is demonstrated that the solution of a Riemann problem consists of a rarefaction *or* a shock wave. With the function $f(\rho)$ represented in Figure 3.18, there exists a Riemann problem where the two waves are *simultaneously* present in the solution. However this lack of concavity does not present any problems

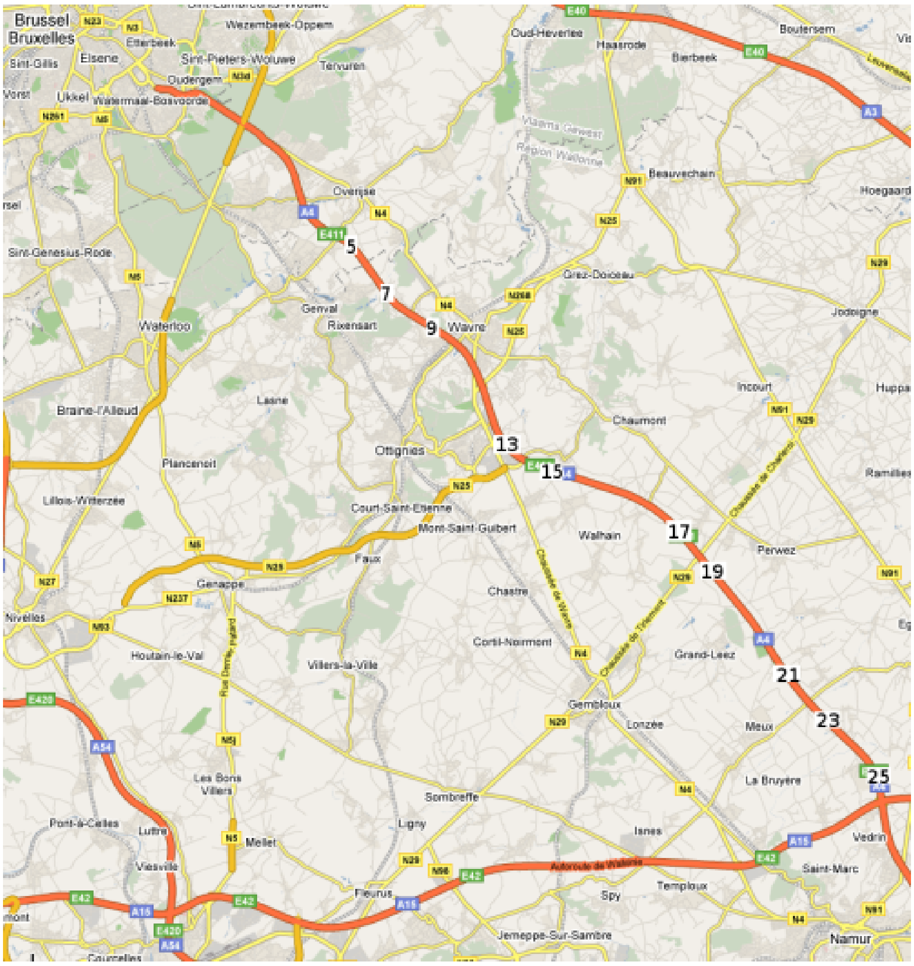


Figure 3.16: The E411 highway between Namur and Brussels and the position of the counting stations.

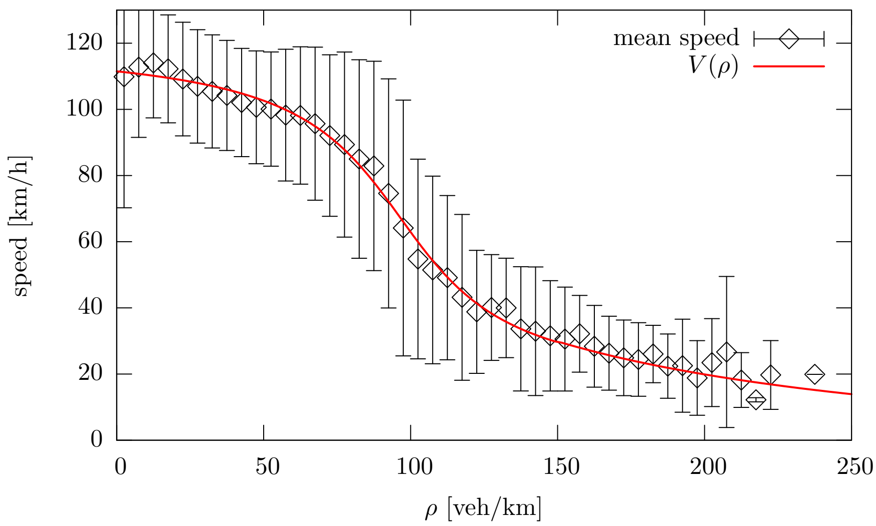


Figure 3.17: The mean speed in function of the density and the 90% interval confidence.

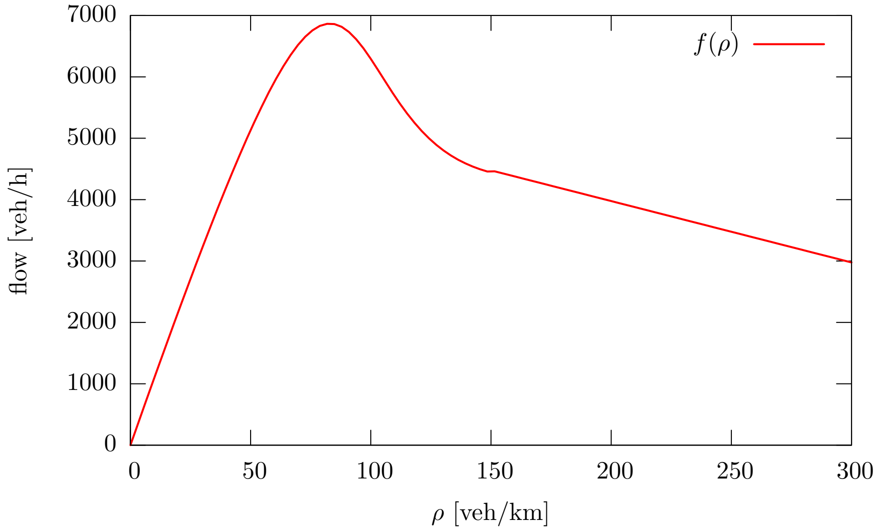


Figure 3.18: The function $f(\rho) = \rho V(\rho)$ based on the measurements performed on the E411 highway.

from the engineering point of view. Indeed, it has no influence on the model assumptions, the junction modelling, the numerical scheme and the control methods presented in Chapter 4. This function can thus be used to validate the numerical simulator.

On the basis of the flows measured after and before the access ramps, the input flows at the on-ramps may be estimated. With these inflows and the function $V(\rho)$ represented in Figure 3.17, a simulation can be performed to predict the evolution of the density. In Figure 3.19 the results of the simulation on the E411 between Daussoulx and Louvranges are represented for October 13, 2003. This day corresponds to a Monday where all the data were available. The simulated density is represented in red and the measured data in blue. Each graph represents the density in function of the time at a particular position. These particular positions correspond to the position of the counting posts (post 25: Daussoulx and post 11: Louvranges). It appears that the simulated density is close to the measured one. The classical backward propagation of the morning traffic jam at the entrance of Brussels can be noticed in the three first sub-figures (the station 5, 7 and 9 are the closest to Brussels).

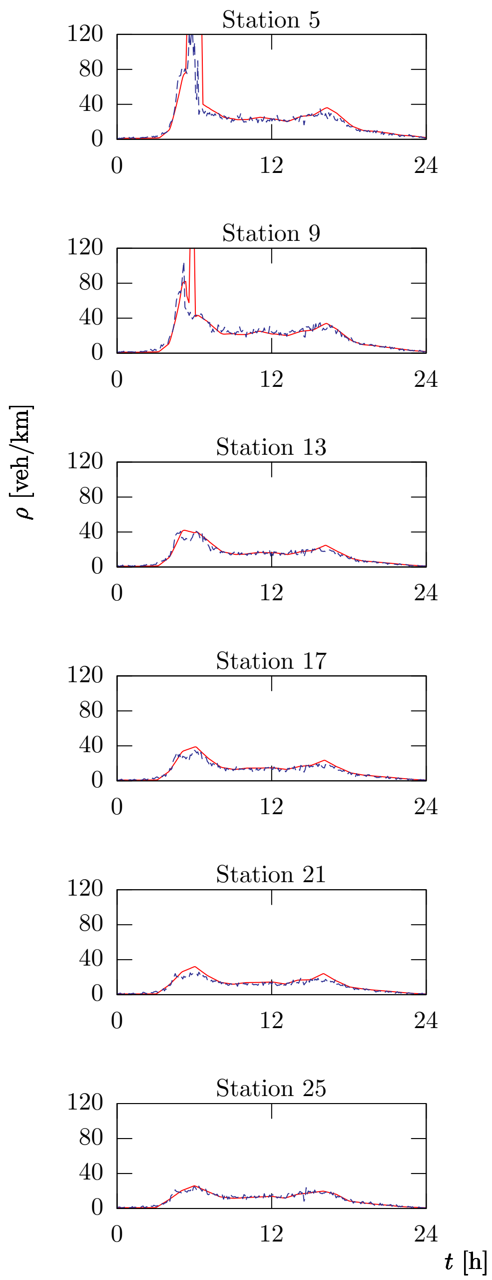


Figure 3.19: The evolution of the traffic density on the E411 on October 13th 2003. The measured densities are the solid red curves and the simulated densities are the dashed blue curves.

Chapter 4

Optimisation and control

With the establishment of a traffic model, it is now possible to consider optimisation and control methods. Optimisation implies the definition of an objective function, i.e. a cost function which needs to be minimised. The simplest function is the Total Time Spent (TTS) by the drivers on the whole network. It is also possible to consider more complex costs including overtime, pollution and accidents. Such costs are developed in Section 4.1. Based on these cost functions optimisation and control methods can be derived.

An important simplification that can be made is to consider steady states, i.e. situations with constant flows. This assumption allows to simplify the problem and to compute the exact solutions taking into account all the complex cost functions derived in Section 4.1. Optimisation performed under steady state assumption allows to obtain for a given network the maximal possible flow or, for a given total flow, the best distribution of the flows between the different paths such that the total cost is minimised. It can also be used to analyse the influence of speed limitations or the building of a new road. Such minimisations are performed in Section 4.2.

The traffic state being rarely constant, other control methods which are not based on a steady state assumption need to be developed. The problem is then more complex and it is usually impossible to find an exact solution with the complex cost functions. For these control methods, simpler cost functions such as the TTS are thus considered. Two methods are proposed in this thesis. The first one (Section 4.3) is a routing method which expresses how to split the flow between two possible paths in order to achieve system-optimality. The second one (Section 4.4) is a feedback ramp-metering control strategy which ensures stability for small perturbations.

4.1 Sustainable cost

The first part of this chapter is dedicated to the derivation of the cost functions. These functions will be used when steady-state optimisations will be considered.

Most of the results presented in this section and in the next one have been produced in the framework of a project for the Science Policy Office from the Belgian State (see [29]).

Sustainability can be defined as an economic state where the people and production demands agree with the preservation of the environment capacity for the future generations ([30]). In this work the issue of sustainability consists in finding the optimal use of the traffic network, in order to reduce the contributions to pollution and noise emission, while satisfying safety and effectiveness for the vehicles. There is an opposition between the user's point of view (reduction of travel time by increasing the speed) and the environment point of view because noise and pollution emissions increase with the speed. In order to analyse the trade-off between these points of view a sustainable cost function is defined which incorporates the econometric costs such as the time cost, the pollution and the accident cost. Of course the relative costs, in monetary units, affected to each of these aspects reflect a political decision.

In section 4.1.1, 4.1.2, 4.1.3 the unit costs are defined for one vehicle, respectively for time, pollution and accident costs, as well as the corresponding costs for the network per hour of operation. The evaluation of the unit costs is based on data provided by KUL/ETE[1]. For the cost evaluation per hour of operation (expressed in € per hour), the steady state situation is first considered: on each arc the traffic conditions are uniform and do not depend on time. For a given flow on a road, there are two possible corresponding densities. The congested state ($\rho > \sigma$) is never optimal. For the uncongested states ($\rho < \sigma$), the speed can be approximated as independent of the density (see Fig. 3.1). This is of course an approximation but not totally irrealistic for low densities and allows a very elegant final formulation (linear in flows). This assumption implies that the speed on each arc is known, either equal to the corresponding $v^{\max}$ or equal to the imposed speed limit, and does not depend on the flow on the arc. Finally the costs are expressed for general non-steady state situation.

In this chapter the following notations are used. A traffic network is described as a directed graph $G = (N, A)$ composed of a set of nodes N connected by a set of arcs A. The nodes (resp. arcs) of the graph represent the junctions (resp. roads) of the network. In addition, a network has some origin nodes ($O \subset N$) and some destinations nodes ($D \subset N$). A path k is a sequence of arcs that are travelled between an origin $r \in O$ and a destination $s \in D$. A graph can possibly have different paths k for the same OD pair. The set of paths between an origin $r \in O$ and a destination $s \in D$ is denoted K_{rs}. The Kronecker delta is used to express if a road belongs to a particular road

$$\delta_{ak} = \begin{cases} 1 & \text{if arc } a \text{ belongs to path } k \\ 0 & \text{otherwise.} \end{cases}$$

Each arc corresponds to a uniform road segment characterised by its length d_a and its speed limit v_a.

[1]Katholieke Universiteit Leuven / Energy, Transport and Environment.

4.1.1 Time cost

The "time cost" includes the direct vehicle operating cost (cost of fuel and additional running costs including repair and maintenance as a factor of fuel cost) as well as the time cost of vehicle occupants.

The definition of unit time cost in monetary value is based on Value of Time units (VOT, [16]), expressed in € per h and per vehicle. The nominal VOT corresponds to the nominal travel time of the vehicle. If the travel time exceeds this nominal value (overtime), the value of the time cost increases to VOT^*, for the part exceeding the nominal time. The VOT depends also on the type of the journey (commuter, business, HGV[2], etc.). In this work an average unit cost is used, computed according to the proportions of the journey types on the Belgian roads. The unit cost for an equivalent vehicle reflecting these proportions, for nominal and overtime, are defined as weighted sums of the VOT:

$$C_{nom} = \sum_{\varrho \in R} \alpha_\varrho (VOT)_\varrho \quad \text{and} \quad C_{ot} = \sum_{\varrho \in R} \alpha_\varrho (VOT)^*_\varrho$$

where R is the set of journey types. The values of α_ρ and (VOT), $(VOT)^*$ are given in Table 4.1, according to the data provided by ETE. The resulting unit cost are given by

$$C_{nom} = 7.61 \left[\frac{€}{veh \cdot h}\right] \qquad C_{ot} = 10.15 \left[\frac{€}{veh \cdot h}\right].$$

	commuting	business	leisure	HGV
α	0.48	0.03	0.44	0.06
VOT	6	21	4	40
VOT^*	9	31.5	4	40

Table 4.1: Time costs: proportion and VOT for different journey types, in € per hour per vehicle, [16].

The travel time along a path k, between origin r and destination s (i.e. $k \in K_{rs}$) is given as

$$t_k = \sum_{a \in A} \delta_{ak} \frac{d_a}{v_a} \tag{4.1}$$

where d_a is the length of arc a, and v_a is the known speed on this arc. In order to take into account the difference of VOT between the nominal and the overtime parts, it is necessary to decompose the travel time into these two components. The nominal travel time along k (no speed limitation) is

$$t_k^{nom} = \sum_{a \in A} \delta_{ak} \frac{d_a}{v_a^{\max}}.$$

[2]Heavy Goods Vehicle: large trucks used for transporting goods.

The overtime due to speed limitation on arc a being

$$\Delta t_a = \frac{d_a}{v_a}\left(1 - \frac{v_a}{v_a^{\max}}\right),$$

it implies that the total overtime due to speed limitation on path k is

$$t_k - t_k^{nom} = \sum_{a \in A} \delta_{ak} \frac{d_a}{v_a}\left(1 - \frac{v_a}{v_a^{\max}}\right) = \sum \delta_{ak} \Delta t_a.$$

This is illustrated in Figure 4.1. Defining the nominal time from r to s as

$$t_{rs}^{nom} = \min_{k \in K_{rs}} t_k^{nom},$$

the overtime due to the choice of path k is

$$\Delta t_k^{nom} = t_k^{nom} - t_{rs}^{nom} \qquad k \in K_{rs}.$$

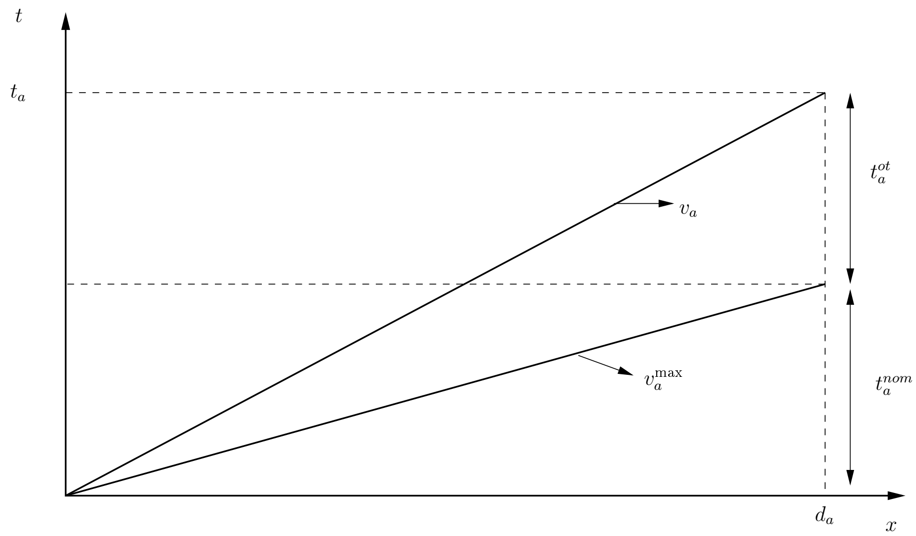


Figure 4.1: Overtime due to speed limit ($v_a < v_a^{\max}$).

With these definitions, the travel time on path k can be expressed as

$$t_k = t_{rs}^{nom} + \Delta t_k^{nom} + \sum_a \delta_{ak} \Delta t_a$$

and the cost of the travel along path k is given by

$$C_k = C_{nom} t_{rs}^{nom} + C_{ot}[\Delta_k^{nom} + \sum_a \delta_{ak} \Delta t_a]\left[\frac{€}{\text{veh}}\right].$$

The global time cost for one hour of network operation (€/h) is thus

$$\begin{aligned} J_T &= \sum_{k \in K} h_k C_k \\ &= \sum_{k \in K} h_k \left\{ C_{nom} t_{rs}^{nom} + C_{ot} \left[\Delta t_k^{nom} + \sum_{a \in A} \delta_{ak} \Delta t_a^{nom} \right] \right\} \quad \left[\frac{€}{\mathrm{h}} \right]. \end{aligned}$$

where h_k is the flow using the path k. It can be seen, that under the constant speed assumptions, this cost is linear in the variables h_k.

4.1.2 Pollution cost

The pollution cost function expresses in monetary value the social cost (diseases and deaths, lost of working days, etc.) which is associated to the pollution gases. The function is based on various parameters such as car type, pollutant type and car age. Emission factors are evaluated for various pollutants (CO_2, CO, NO_x, benzene, hydrocarbon, butadiene 1,3 and dust) by vehicles type. Vehicle categories are petrol and diesel cars, LGV, HGV, buses and their proportion according to the Belgian fleet. Ages of these vehicles is also taking into account with the EU standards (pre-Euro to Euro IV). Finally, each pollutant has a cost associated to a roadway type (highway, secondary road and city roads).

For each pollutant and each type of vehicle, there is an emission factor giving the mass of pollutant emitted for 1 km travel by this type of vehicle. This factor is a function of the speed. Figure 4.2 provides the typical examples of emission factors (expressed in g/km), in function of speed, for various gases and vehicles. Most of the pollutant present high values under 30 km/h and over 90 km/h. Thus low speed associated to congestion effect contributes significantly to pollutant emission. Trucks and buses exhibit the highest pollution emission. PM emission represents the dust emission.

For the free flow conditions, when the speed is known independently of the traffic conditions, the pollution cost per km on a given arc depends therefore on the various emission factors, on the proportion of vehicles types and on the economical cost associated to each pollution according to the type of road (highway, secondary or city roads). It is expressed as

$$C_a^P = \sum_{\sigma \in S} \alpha_\sigma \left[\sum_{p \in \Pi} C_p(e_a) F_{p\sigma}(u_a) \right] \qquad \left[\frac{€}{\mathrm{veh\ km}} \right] \tag{4.2}$$

In this expression

- Π is the set of pollutant types;
- S is the set of vehicles types;
- α_σ is the proportion of vehicles of type σ;

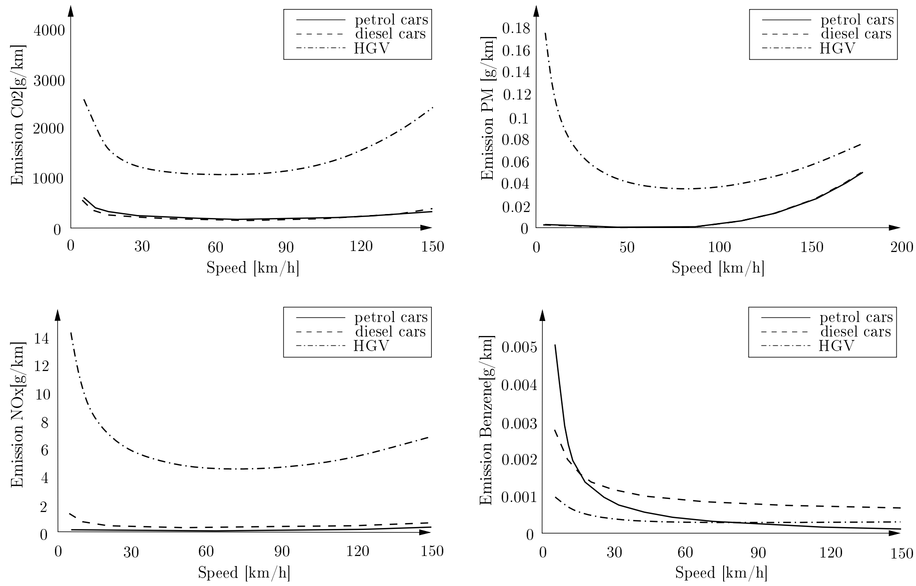


Figure 4.2: Pollutant emission factors, value (g/km) in function of speed for different vehicles types (cars with 2.0 l, Euro IV, except for CO_2 Euro II).

- $F_{p\sigma}(u_a)$ is the emission factor for pollutant p and vehicle σ in g per vehicle and per km (see Fig. 4.2);
- e_a is the road type of arc a;
- $C_p(e_a)$ is the cost of pollutant p (in €/g) for an arc of type e_a.

The values of the variables involved in this expression are deduced from data provided by KUL/ETE:

- the functions $F_{p\sigma}(u_a)$ see Fig. 4.2;
- the proportions of vehicles α_σ: 35% petrol cars, 35% diesel cars, 30% HGV;
- the costs associated to each pollutant and each road type $C_p(e_a)$.

The resulting values of C_a^P, in function of the speed, for each road type are given in Fig. 4.3. The particular values corresponding to the speeds considered under steady-state traffic are summarised in Table 4.2.

For a given arc a of length d_a, the pollution cost associated to an equivalent vehicle is equal to $C_a^P d_a$. For one hour of operation, corresponding to a flow of q_a vehicles on the arc, and for the whole network, the global pollution cost is given by

$$J_P = \sum_{a \in A} C_a^P \ q_a \ d_a \qquad [€/\text{h}].$$

	120 [km/h]	90 [km/h]	70 [km/h]
highway	$2.3\,10^{-2}$	$1.55\,10^{-2}$	$1.42\,10^{-2}$
secondary		$0.83\ \ 10^{-2}$	$0.72\,10^{-2}$
city			$1.82\ \ 10^{-2}$

Table 4.2: Values (in $\frac{€}{veh \cdot km}$) of C_a^P for different road types and for different velocities.

It must be pointed out that, under an assumption concerning the speed on each arc, this cost is linear with respect to the flows q_A on the different arcs.

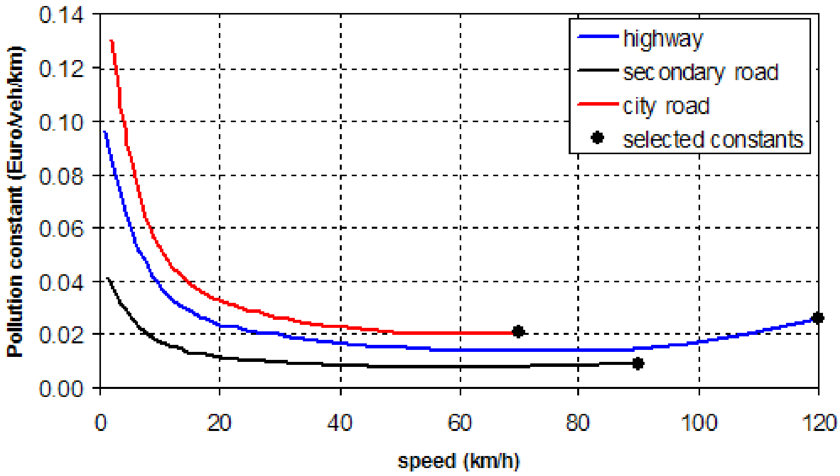


Figure 4.3: Pollution constants value in function of speed for different road types.

4.1.3 Accident cost

Accident costs reflect the cost of collision between vehicles according to an accident risk. Costs are related to the willingness of users to pay for security and the part taken by the society.

The risk is defined in function of the number of accident between vehicles categories and the annual travelled distance-vehicle. Risk calculation depends also on vehicle speed v_a and on the type of accident *severity* (fatal, severe, and light). Vehicle types represented are car, bus and HGV. Based on this statistical risk evaluation, an accident cost constant C^A can be evaluated depending on the speed v_a on arc a; according to the data provided by KUL/ETE

$$C^A(v_a) = 1.665\,10^{-3} v_a^2 + 1.7\,10^{-5} v_a^3 \qquad [€/\text{veh km}].$$

As for the pollution cost, the accident cost on the network is computed as

$$J_A = \sum_{a \in A} C^A(v_a) q_a d_a. \tag{4.3}$$

4.1.4 Global cost

The global cost function includes the three components presented in the previous sections:

$$J_G = J_T + J_P + J_A.$$

This global cost is a linear combination of the variables h_k and q_a. The flow on arcs q_a can also be expressed as a linear combination of the path flow h_k. The global cost is thus linear with respect to h_k. This is the reason why the approximation of a constant speed on a arc was used.

If the approximation of a speed independent of the flow is too important or if non steady-state is considered, a more general formulation must be used for the costs. The time cost rate for the whole network is then defined as

$$\frac{dJ_T}{dt} = \sum_{a \in A} \int_0^{d_a} C_a(\rho_a(x,t), v_a(x,t)) dx \tag{4.4}$$

where

$$C_a(\rho_a, v_a) = C_{nom} \rho_a(x,t) \left(\frac{v_a(x,t)}{v_a^{\max}} \right) + C_{ot} \rho_a(x,t) \left(1 - \frac{v_a(x,t)}{v_a^{\max}} \right).$$

The pollution cost rate is

$$\frac{dJ_P}{dt} = \sum_{a \in A} \int_0^{d_a} C_a^P(v_a(x,t)) v_a(x,t) \rho_a(x,t) dx$$

where the unit cost $C_a^P(v_a)$ represents the pollution cost per covered kilometer, and therefore leads to infinity for v_a tending to zero. The product $(C_a^P(v_a)v_a)$ is the pollution cost per unit time (hour). This quantity tends to a finite value, for v_a tending to zero.

The accident cost rate is

$$\frac{dJ_A}{dt} = \sum_{a \in A} \int_0^{d_a} \left[C^A(v_a(x,t)) v_a(x,t) \rho_a(x,t) \right] dx.$$

The global cost rate J_G is obtained by integrating the different cost rates

$$J_G[0,T] = \int_0^T \left(\frac{dJ_T}{dt} + \frac{dJ_P}{dt} + \frac{dJ_A}{dt} \right) dt. \tag{4.5}$$

4.2 Steady-state optimisation

When considering steady-states, two main questions must be addressed: for a given network what is the maximal possible flow and, for a given flow, what is the best distribution of drivers between the different possible paths in order to minimise the cost function?

These questions will be illustrated with the network represented in Figure 4.4. There are three origins (O1, O2 and O3) and three destinations (D1, D2 and D3). The Upper branch, U represents a highway with three lanes and with a maximum speed of 120 km/h. The Middle branch, M represents a secondary road of two lanes with a maximum speed of 90 km/h. The Lower branch L is a single lane road with a maximum speed of 70km/h. Four road segments connect them together (ac, ef, gh, jk). To represent the fact that the drivers have fixed destination aims, an Origin–Destination (OD) matrix is introduced (see Table 4.3). The element at row r and columnm s of the OD matrix represents the fraction of the drivers entering at O_r who are going to D_s. The lengths of the road segments represented by the graph are given in Table 4.5. The different paths between the origins and the destinations are given in Table 4.6. They are numbered from 1 to 18 and have also a colour code shown in Tab. 4.4

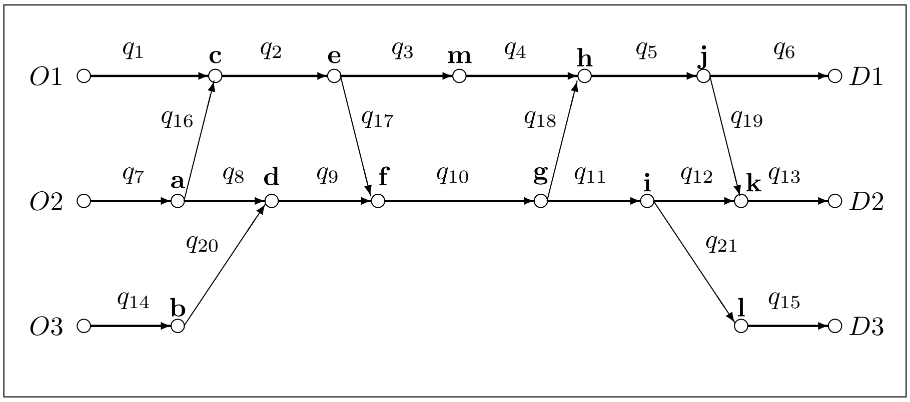


Figure 4.4: The benchmark network

	D1	D2	D3
O1	70%	15%	15%
O2	15%	70%	15%
O3	50%	50%	-

Table 4.3: Origin–Destination matrix, m_{rs}.

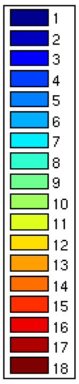


Table 4.4: The colour code of the different paths

arc	d_a	road type	arc	d_a	road type
#	km		#	km	
1	2.3	highway	11	1.2	national
2	1.1	highway	12	0.9	national
3	2.3	highway	13	1.8	national
4	2.3	highway	14	1.8	city
5	1.1	highway	15	1.8	city
6	2.3	highway	16	0.5	national-highway $\Rightarrow$ national
7	1.8	national	17	0.5	highway-national $\Rightarrow$ national
8	0.9	national	18	0.5	national-highway $\Rightarrow$ national
9	1.2	national	19	0.5	highway-national $\Rightarrow$ national
10	3.6	national	20	0.5	city-national $\Rightarrow$ city
			21	0.5	national-city $\Rightarrow$ city

Table 4.5: Lengths of road segments.

path	arc	origin	destination	path length
k	a	r	s	km
1	1-2-3-4-5-6	1	1	11.4
2	1-2-3-4-5-19-13	1	2	11.6
3	1-2-17-10-21-15	1	3	11.7
4	7-16-3-4-5-6	2	1	11.6
5	7-16-2-3-4-5-19-13	2	2	11.8
6	7-8-9-10-11-21-15	2	3	11.5
7	14-20-9-10-18-5-6	3	1	11.7
8	14-20-9-10-11-12-13	3	2	11.5
9	1-2-17-10-18-4-5-6	1	1	11.8
10	1-2-17-10-11-12-13	1	2	11.6
11	7-8-9-10-18-5-6	2	1	11.6
12	7-8-9-10-11-12-13	2	2	11.4
13	7-16-2-17-10-18-5-6	2	1	12.0
14	7-16-2-17-10-11-12-13	2	2	11.8
15	7-8-9-10-18-5-19-13	2	2	11.8
16	7-16-2-17-10-18-5-19-13	2	2	12.2
17	7-16-2-17-10-11-21-15	2	3	11.9
18	14-20-9-10-18-5-19-13	3	2	11.9

Table 4.6: Paths between origins and destinations.

4.2.1 Flow maximisation

In this section the problem of flow maximisation is addressed. The main question is "What is the network capacity". More precisely, the question is to compute, for a given network and given drivers' intentions, the maximal flow able to pass through the network. This procedure allows to analyse the influence on the network capacity of the construction of a new road or the setting of a new speed limitation.

The drivers' intentions are represented by the OD matrix M (see [13]). This matrix is defined such that an element m_{rs} gives the proportion of the flow entering at origin r that exits at destination s. It implies

$$\sum_{s \in D} m_{rs} = 1 \qquad \forall r \in O.$$

Table 4.3 is an example of such an OD matrix. Let v_r denote the fraction of the total flow through the network originating from r. One has

$$\sum_{r \in O} v_r = 1.$$

The steady-state flow on a path k is denoted h_k. The maximum flow problem is formulated as:

$$\max_{h_k} \qquad \sum_{r \in O} \sum_{s \in D} \sum_{k \in K_{rs}} h_k \tag{4.6}$$

subject to

$$\text{arc flow capacity} \qquad \sum_{r \in O} \sum_{s \in D} \sum_{k \in K_{rs}} \delta_{ak} h_k \leq f_a^{\max} \qquad \forall a \in A \tag{4.7}$$

$$\text{non negative flows} \qquad h_k \geq 0 \qquad \forall k \in K \tag{4.8}$$

where $f_a^{\max}$ is the maximal possible flow on arc a. The left part of the first constraint (4.7) represents the relation between the path flow h_k and the arc flows $f_a^{\max}$. The flow on an arc a, must be smaller than its capacity $f_a^{\max}$.

In order to satisfy the OD matrix, the objective function (4.6) may also be subjected to

- satisfaction of O-D matrix

$$\underbrace{\sum_{k \in K_{r,s}} h_k}_{\text{flow from } r \text{ to } s} = m_{rs} \underbrace{\sum_{x \in D} \sum_{k \in K_{rx}} h_k}_{\text{flow from } r} \qquad \forall r \in O, \forall s \in D \tag{4.9}$$

- distribution among the origins

$$\underbrace{\sum_{s \in D} \sum_{k \in K_{rs}} h_k}_{\text{flow from } r} = v_r \underbrace{\sum_{r \in O} \sum_{s \in D} \sum_{k \in K_{rs}} h_k}_{\text{total flow}} \qquad \forall r \in O \tag{4.10}$$

arc	case 1	case 2	case 3	case 4
#	speed (km/h)			
1	120	120	120	120
2	120	120	90	90
3	120	90	90	70
4	120	90	90	70
5	120	120	90	90
6	120	120	120	120

Table 4.7: Speed limitations on the highway.

This is a linear optimisation problem and can thus be very efficiently solved using classical algorithms. The size of the problems under consideration being small (1 variable per path), an algorithm such as the *simplex* method provides instantaneous results. This algorithm is implemented in most of the scientific toolboxes (function `linprog` in MATLAB).

This problem of flow maximisation can be illustrated using the benchmark network represented in Figure 4.4. The problem is considered with the constraints (4.7), (4.8), (4.9) but *without* the constraint (4.10). With the OD matrix constraint (4.9), the distribution of flows among the arcs is represented in Fig.4.5 (case 1) and the maximum input is 9793 veh/h (see Tab. 4.8). Arcs 2,5 and 10 are saturated.

In order to evaluate the influence of speed limitations on the maximum flow, the three additional situations presented in Table 4.7 are considered. More precisely, the following speed limitations on the highway are considered:

- case 2 : 90 km/h on arcs 3 and 4;
- case 3 : 90 km/h on arcs 2,3,4,5;
- case 4 : 90 km/h on arcs 2,5 and 70 km/h on arcs 3,4.

The results are given in Table 4.8 and illustrated in Fig. 4.5. It can be observed that the speed limitations induce only a small decrease of the maximum flow (from 9793 veh/h to 9299 veh/h) although the speed limitations are rather significant. In Fig. 4.6, the distribution of the flow among the various possible paths is represented. It is observed that, for each OD pair, almost all the traffic takes the most natural and direct path. However, one can see that the flow from origin 2 to destination 2 (city-city) must be split between the highway and the national road in order to achieve the maximum flow.

Another interesting situation may be investigated: the case where all the drivers use only the shortest path (in time) between an origin and a destination. In this case, the maximum flow becomes 6990 veh/h (case 1 in Table 4.8), which is significantly less than the optimal solution. One important observation is that the total flow may be higher if speed limitation is used (7713 veh/h for case 4). The reason can be seen in Fig. 4.7 and 4.8: with the speed limitations the paths passing by the highway need more time and some drivers have to change their path between the origin and the destination. The particular case of the drivers going from origin 3 to destination 2 can be considered: in the first case, they take the path 18 passing through arc 5 which is saturated; in case of speed limitation, the drivers use path 8 leaving some flow for arc 5 available for other drivers. It can be seen here how speed limitations can be used as a tool for increasing the flow through a road traffic network. The extreme case where the suppression of a road leads to a higher network capacity is known as the Braess' paradox and is illustrated in Section 4.3.2.

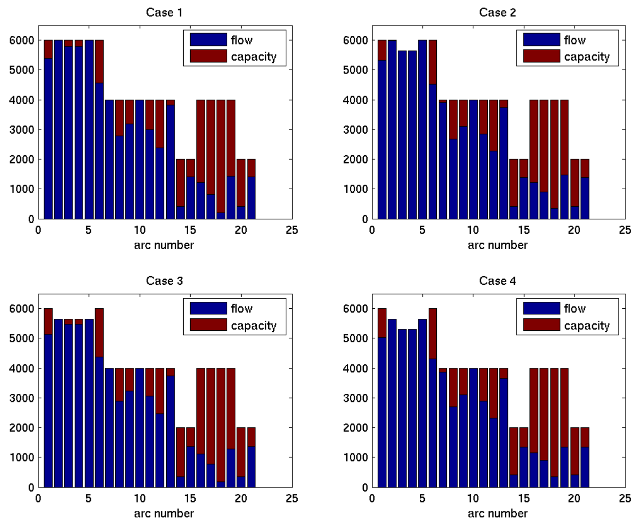


Figure 4.5: Max flow results for the benchmark network: the flows on the different arcs.

Case	all paths	paths shortest in time
#	max flow (veh/km)	
1	9793	6990
2	9651	6983
3	9480	7472
4	9299	7713

Table 4.8: The maximal flow for the benchmark network

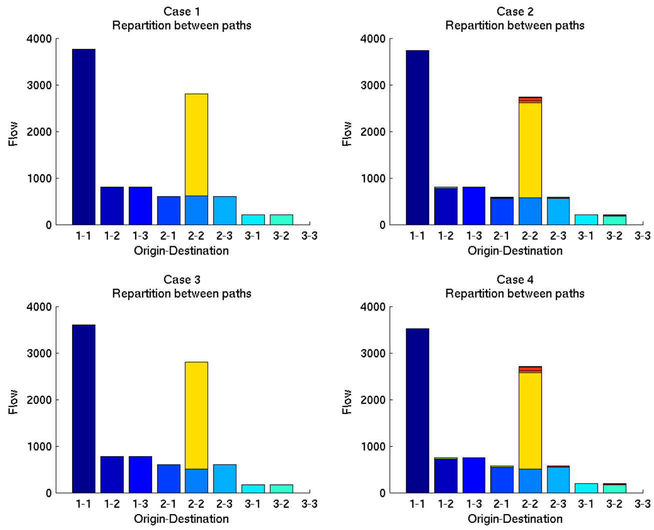


Figure 4.6: Max flow results for the benchmark network: distribution among the different paths.

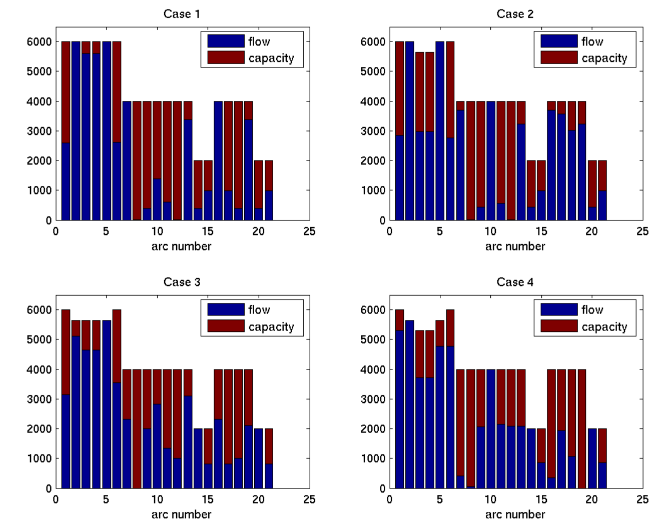


Figure 4.7: Max flow results for the benchmark network: the flow on the different arcs where the shortest path in time are used by all the drivers.

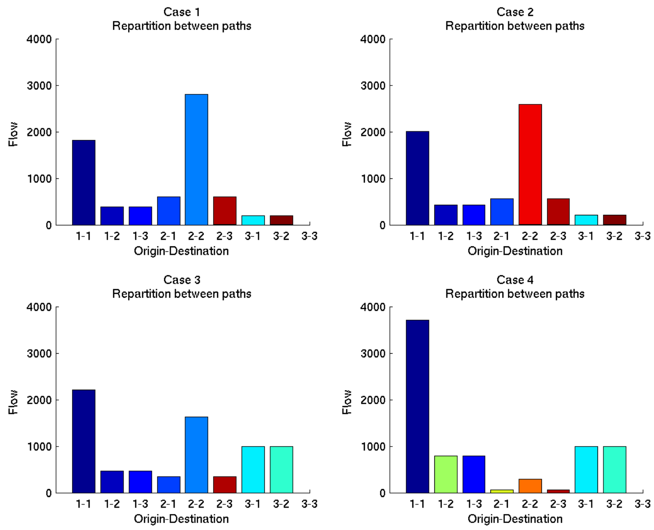


Figure 4.8: Max flow results for the benchmark network: the distribution between the different paths. The shortest path in time is used by the drivers.

4.2.2 Sustainable cost minimisation

In this section the problem of the optimal distribution of the traffic flows among all possible paths for a given demand is addressed. More precisely, the question is to find, for given inflows and drivers' intentions (the OD matrix), the best flow repartition in order to minimise the total cost J_G defined in section 4.1. Obviously the total inflow imposed F_{in} must be less than the network capacity otherwise there is no solution. Since steady-states are considered, the speeds on the different arcs can be considered as given constant parameters of the problem (either a speed limit v_a or the maximal speed $v_a^{\max}$).

The minimisation problem is described as follows:

$$\min_{h_k} J_G = \min_{h_k} \big(J_T + J_P + J_A\big) \tag{4.11}$$

subject to

- a given demand
$$\sum_{r\in O}\sum_{s\in D}\sum_{k\in K_{rs}} h_k = F_{in}$$
- the requirements of O–D matrix
$$\sum_{k\in K_{r,s}} h_k = m_{rs}\sum_{x\in D}\sum_{k\in K_{rx}} h_k \qquad \forall r\in O, \forall s\in D$$
- the distribution trough origins
$$\sum_{s\in D}\sum_{k\in K_{rs}} h_k = v_r\sum_{r\in O}\sum_{s\in D}\sum_{k\in K_{rs}} h_k \qquad \forall r\in O.$$

As in Section 4.2.1, it is a linear problem which can be very efficiently solved using a classical linear programming algorithm.

This problem is illustrated with the benchmark network (see Fig. 4.4). The optimisation results corresponding to various conditions are summarised hereafter:

1. As in Section 4.2.1, four cases are considered corresponding to different speed limitations (see Table 4.7).
2. The total demand is fixed at 7000 veh/h (4000 from origin O_1, 2000 from O_2 and 1000 from origin 0_3), with the m_{rs} matrix represented in Table 4.3.

   It must be pointed out that, for case 1 and case 2, the global flow exceeds the maximum flow achievable if each user follows its shortest path (see Table 4.8). Several paths between each O-D pair are therefore necessary in order to satisfy the demand. On the other hand, this demand is less than the maximum possible flow, this implies that the network is not saturated and several degrees of freedom are available in order to minimise the cost criterion.

3. The optimisations are performed using 3 types of criterion : minimisation with respect to the global cost J_G, with respect to the time cost J_T or to the pollution cost J_P.

The results are summarised in Table 4.9 giving, for each minimisation criterion and for each case, the values of the different contributions to the cost (time, pollution and accident) and the value of the global cost. It can be observed that the time cost component is the most determinant and that the accident cost is very low. As expected the effect of speed limitations is to increase the time cost and to decrease the pollution and accident costs.

Minimisation with respect to the global cost or to the time cost only produces almost the same values for the cost value, and mostly the same distribution between the possible paths. The reason is that the time cost represents the major part of the global cost. If the minimisation is performed with respect to the pollution cost, the pollution cost decreases with, as a result, an augmentation of the time cost and of the global cost.

It is possible to represent the flows on the different arcs and on the different paths. Figure 4.9 and 4.10 give these results for the four cases, when the optimisation is performed with respect to the global cost. In case 1, saturation occurs on the highway (arc 5) while for the 3 other cases saturation occurs on arc 10. The effect of the speed limitations on the distribution between the possible paths for all O-D pairs is represented on Fig.4.10. Consider for instance the (2-2) pair. In case 1 all users follow path 5, for case 2 most follow path 14, while in case 2 and 4 most use path 12.

Figure 4.11 represents the flows on the arc, when the optimisation is performed with respect to the pollution cost. The flow distribution can be compared with Fig. 4.9. In this situation, saturation occurs on the same arc (arc 5) for the four cases. In fact the flow on the highway decreases with augmentation of the middle branch flow. In order to minimise pollution the arcs with smaller velocity are privileged and more drivers avoid the highway, accepting some travel time augmentation.

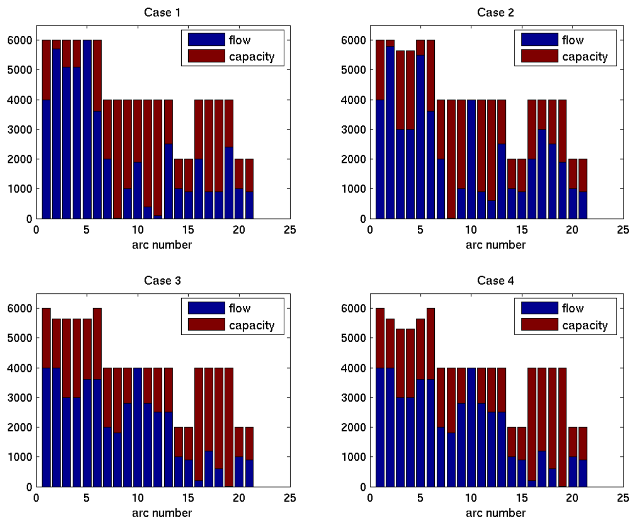


Figure 4.9: Vehicle flow on arcs, optimisation with respect to the global cost.

criterion	case	Global Cost	Time Cost		Pollution Cost		Accident Cost	
		[€/h]	[€/h]	%	[€/h]	%	[€/h]	%
Global cost	1	6793	5171	76	1491	22	131	2
	2	7603	6246	82	1236	16	121	2
	3	8003	6778	85	1111	14	115	1
	4	8870	7667	86	1093	12	110	1
Time cost	1	6793	5171	76	1491	22	131	2
	2	7603	6246	82	1236	16	121	2
	3	8003	6778	85	1111	14	115	1
	4	8870	7667	86	1093	12	110	1
Pollution cost	1	6995	5583	80	1281	18	131	2
	2	7657	6361	83	1175	15	121	2
	3	8003	6778	85	1111	14	115	1
	4	8870	7667	86	1093	12	110	1

Table 4.9: The different costs for the benchmark network.

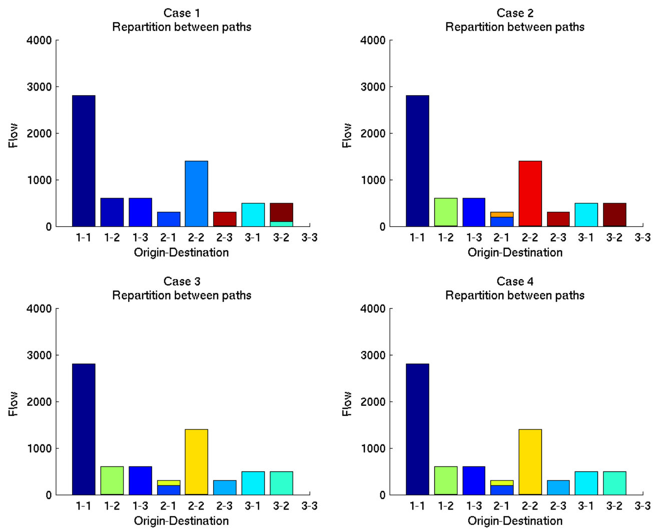


Figure 4.10: Distribution of the flows between the possible paths, optimisation with respect to the global cost.

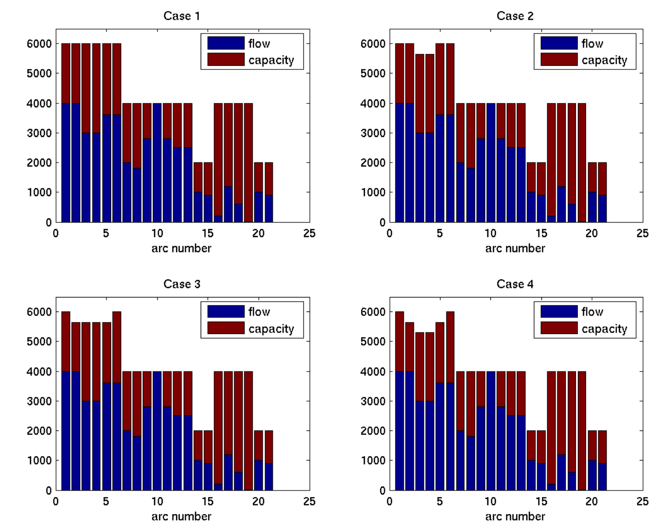


Figure 4.11: Vehicle flow on arcs, optimisation with respect to the pollution cost.

4.3 System-Optimal routing

Steady-state optimisation is not always the proper way to improve the efficiency of an existing network. Such an optimisation is useless if the inflows are unknown and variable. If an accident occurs the transient effects are important and not taken into account in a steady-state representation. It is thus important to develop control methods which take into account the full dynamics of the system. Since this problem is much more complex than the steady-state case, simpler cost functions such as the TTS are used. The first method developed here is a “routing strategy”.

Usually there are different paths available for a driver to reach a particular destination from a particular origin. One can think of the ring around a city where there are two possible paths between two particular ramps. It can be assumed that it is possible, for a network operator, to influence the fraction of the drivers taking one of the two paths, for example by presenting some informations about the time needed on each path. The problem of “optimal routing” consists for the operator to adjust the “splitting rate” at some bifurcation node in order to optimise a cost criterion representing the efficiency of the network. When real-time route guidance is considered, a perfect solution is usually not achievable since the optimal splitting rates depend on the value of the inflows for future time which are not available in real time situations. It is this real-time case which is addressed by the new routing strategy presented in this section.

A lot of routing algorithms have already be developed using network models where it is assumed that the flow is independent of the position on the road. It means that the evolution of the traffic on a single road is not modelled. In these works, the time spent by the drivers on this road can be a function of the flow (see e.g. [37]) or a random variable with time-dependent distribution (see [18] for a review). Fewer works have been done using a network model, like the LWR model, where the evolution of the traffic state on each road is described. One of these works [59] is presented and improved.

4.3.1 An existing User–Optimal strategy

The routing strategy presented in [59] is focused on User–Optimality (UO). An user–optimum or user–equilibrium situation occurs when *each driver* chooses the route minimising his own travel time. If a driver had complete information about the traffic state, it would be his normal behaviour. Even without omniscience this situation tends to occur if the traffic demand remains constant and there is no accident. In that case, day after day, the drivers try new paths, obtain new information and finally select the best path. When such situation is achieved it is an equilibrium since, if nothing external happens, none of the drivers has an advantage to modify his route. This corresponds to a Nash Equilibrium in game theory.

For a bifurcation node where two outgoing paths may be used to reach a final destination the UO conditions may easily be expressed. The splitting rate α expresses the fraction of the flow leaving the node using the first outgoing path; if the time spent by the drivers to reach the final destination using the outgoing path i is denoted by τ_i then the UO condition are:

$$\begin{aligned} \Delta\tau(t) \geq 0 \quad &\text{if} \quad \alpha(t) = 1 \\ \Delta\tau(t) = 0 \quad &\text{if} \quad 0 < \alpha(t) < 1 \\ \Delta\tau(t) \leq 0 \quad &\text{if} \quad \alpha(t) = 0 \end{aligned} \tag{4.12}$$

where $\Delta\tau(t) = \tau_2(t) - \tau_1(t)$. This condition expresses the fact that, as long as the time used to travel on one of the two paths is lower than the time on the other path, the drivers will choose the first one.

This UO situation is usually not achieved (variable inflows, accidents,...). The idea behind the routing strategy presented in [59] is to tend to the UO by adapting the splitting rate α. More precisely by keeping $\Delta\tau$ as close as possible to zero if α is between zero and one. The network operator can apply this strategy by adapting the factor α using a feedback controller:

$$\alpha(t) = K_p\Delta\tau(t) + \int K_i\Delta\tau(t) \tag{4.13}$$

where $\alpha(t)$ is clipped between zero and one. In general, it is difficult to use the exact value of $\Delta\tau(t)$ since this value depends on the traffic state for time greater than t and thus the inputs of the network must be known in advance. One can use the *instantaneous travel time*, the time spent by the driver assuming that the traffic state will not change, to approximate $\Delta\tau(t)$. This procedure to adapt the splitting rate α has shown to improve the global efficiency of the network. For example, if an accident occurs on a path this strategy will send most of the drivers on the other path. Nevertheless this strategy can still be improved by using a System–Optimal (SO) criterion instead of a User–Optimal criterion.

4.3.2 Difference between User–Optimality and System–Optimality

There is a difference between User–Optimality and System–Optimality. A SO situation corresponds to a particular choice of the splitting rate α which minimises a cost function. A usual simple choice made for the cost is the *Total Time Spent*:

$$\text{TTS} = \iint \rho(x,t)\ dxdt \tag{4.14}$$

where the integration is done over the whole network for all time. This cost is a "social cost" which represents the time spent by all the drivers on the network. With this definition, the cost rate $M(t)$ is

$$M(t) = \int \rho(x,t)dx$$

where the integration is done over each road. Let $M_i(t)$ denotes the contribution of road i to the cost rate

$$M_i(t) = \int_0^{L_i} \rho_i(x,t)dx$$

where L_i is the length of road i.

To illustrate the difference between the user-equilibrium and the system-optimality, the network represented in Fig 4.12 may be considered. In this network there are two classes of drivers. The drivers from the first (resp. second) class enter the network using the first (resp. second) input node and leave it using the first output (resp. second) node. The drivers from the first class may thus choose to use two different paths (the upper and the lower road) while the drivers from the second class can only choose one path (lower road). The factor α represents the fraction of the first drivers using the lower road.

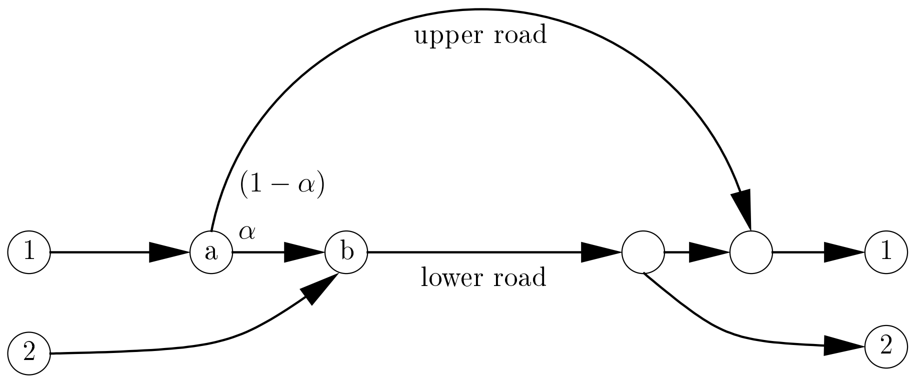


Figure 4.12: A simple benchmark network

In this simple network if the length of the upper road is greater than the length of the lower road then the difference of cost between the user-optimum and the system-optimum can become important. For simplicity, consider that the length of the lower road is 1 km, the length of the upper road is l, the flow at the two inputs is 3000 [veh/h] and the function relating the density to the speed is

$$V(\rho) = 0.74(180 - \rho)[km/h].$$

The cost rates $M(t)$ and the splitting rates corresponding to the user-optimum and the system-optimum situations at the equilibrium are represented in Figure 4.13. It can be seen that the difference between the two costs may be significant, especially for $l = 2$ where the difference is around 30 percent.

This difference between UO and SO explains why it is sometimes preferable *not* to build additional roads. For example consider the small road between the junctions a and b. If this road was not built the drivers from the first class could only take the upper road ($\alpha = 0$). In Figure 4.14 the costs at steady

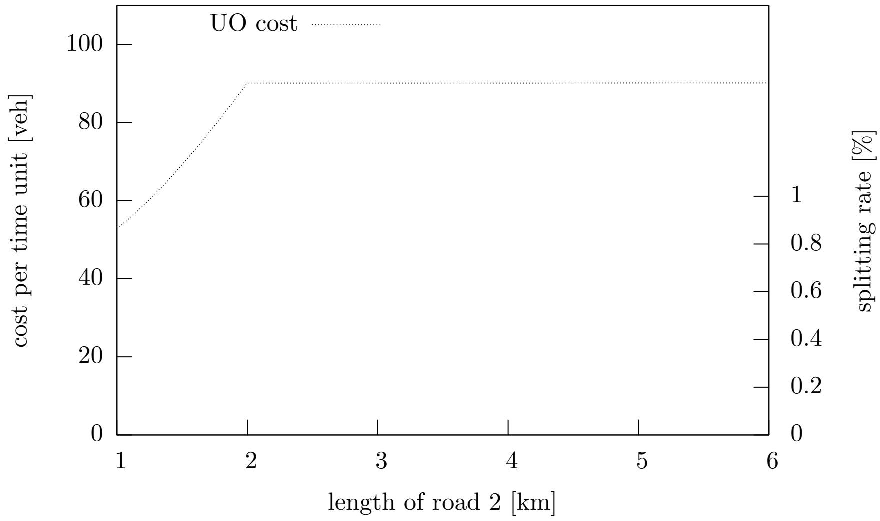


Figure 4.13: The evolution of the user–optimum (UO) and system–optimum (SO) costs and the splitting rates α in function of the length of the upper road.

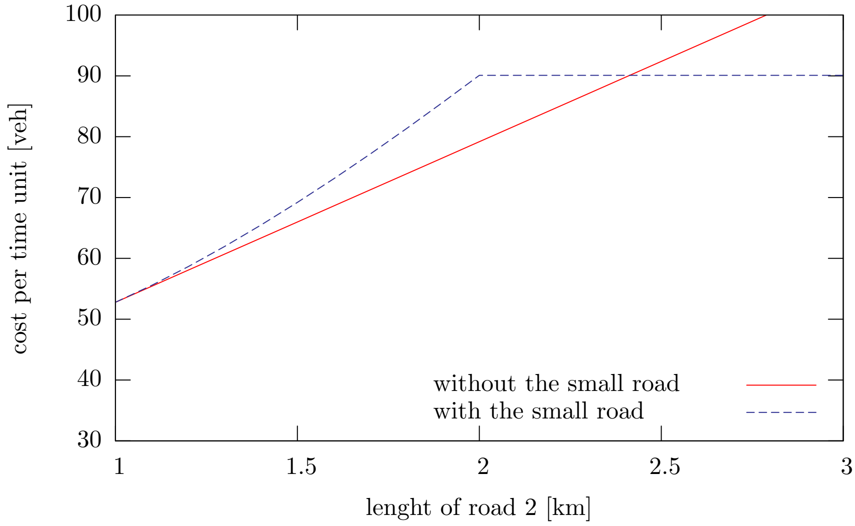


Figure 4.14: The costs at steady state without control (i.e. assuming UO situation) with and without the small road allowing the drivers from the first class to take the lower road.

state without control (i.e. assuming UO situation) with and without this small road are represented. It can be noticed that if the length of the upper road is between 1 and 2.4 km, it is better *not* to build the road.

This non-intuitive conclusion is known as the Braess' paradox (see [53]). It has been experimentaly observed in 1990 when the 42nd Street in New York was closed (see [39]) and in Stuttgart downtown when a new road was added (see [4]).

4.3.3 A new System–Optimal strategy

With this difference between UO and SO in mind, it is possible to improve the strategy (4.13) by analysing the influence of a change of α on the TTS. A change of the splitting rate α by $\Delta\alpha$ implies a change of the inflow of road i by Δq_i. If the density on the road i is strictly less than σ then the density on road i will change by

$$\Delta\rho_i(x,t) \approx \frac{\Delta q_i}{f'(\rho_i(x,t))}.$$

This is an approximation not only because a first order approximation is used but also because it is implicitly assumed that the only change of flow on the path i comes from the change of the splitting rate α. The influence of Δq_i on M_i is then

$$\Delta M_i(t) \approx \int_0^{L_i} \frac{\Delta q_i}{f'(\rho_i(x,t))} dx.$$

Since $\Delta q_1 = -\Delta q_2$, the cost rate total change for the whole network is

$$\Delta M(t) \quad \approx \quad \Delta\alpha q_{in} \left(\int_0^{L_1} \frac{1}{f'(\rho_1(x,t))} \quad - \quad \int_0^{L_2} \frac{1}{f'(\rho_2(x,t))} \right) dx \tag{4.15}$$

where q_{in} is the inflow at the entrance of the bifurcation node. This equation is only valid if $\rho_i(x,t) < \sigma$ since, when this condition is not satisfied at a particular position x, an increase of the inflow of the road i does *not* lead to an increase of the density at x. In this case (4.15) tends to suggest that an increase of the inflow produces a decrease of the TTS (since $f'(\rho_i(x,t)) < 0$ when $\rho_i(x,t) > \sigma$) even it is clearly not the case. In order to alleviate this limitation and to penalise the inflow increase of a jammed road, (4.15) may be modified as

$$\Delta GM(t) \approx \Delta\alpha q_{in} \left(\int_0^{L_1} \frac{1}{g(f'(\rho_1(x,t)))} - \int_0^{L_2} \frac{1}{g(f'(\rho_2(x,t)))} \right) dx \tag{4.16}$$

where

$$g(x) = \begin{cases} x & \text{if } x \geq f'(\sigma-\epsilon) \\ f'(\sigma-\epsilon)\left(2 - \frac{x}{f'(\sigma-\epsilon)}\right)^{-1} & \text{if } x \leq f'(\sigma-\epsilon) \end{cases} \tag{4.17}$$

where $\epsilon > 0$. With this choice of g, ΔGM is equal to ΔM for ρ less than $\sigma - \epsilon$ and an increase of flow on a jammed road is heavily penalised.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
User–optimum strategy	847	1246	754	864
System–optimum strategy	697	731	644	738
Benefit	17.8 %	41.3 %	14.6 %	14.6 %

Table 4.10: The Total Time Spent (veh·h) on the network for the four scenarii.

Based on this particular measure of the TTS evolution, some "system-optimality" conditions may be derived. These conditions are exact with respect to the cost function (4.14) only at equilibrium and if $\rho_i < \sigma$ on all roads due to the various approximations made. Expressed in the same form as (4.12), the conditions are :

$$\begin{array}{lll} \Delta c \geq 0 & \text{if} & \alpha = 1 \\ \Delta c = 0 & \text{if} & 0 < \alpha < 1 \\ \Delta c \leq 0 & \text{if} & \alpha = 0 \end{array} \tag{4.18}$$

where $\Delta c = c_2 - c_1$ and

$$c_i = \int_0^{L_i} \frac{1}{g(f'(\rho_i(x,t)))} dx.$$

These conditions express the fact that while the TTS can be decreased by modifying α, the splitting rate must decrease or increase depending on the sign of Δc. When $\Delta c = 0$, the splitting rate α may be kept constant.

Based on these conditions, and by analogy with (4.13), a new strategy can be derived. The idea of this new strategy is to keep Δc close to zero if α is between zero and one by adapting α using a Proportional-Integral controller:

$$\alpha(t) = K_p \Delta c(t) + \int K_i \Delta c(t) \tag{4.19}$$

where $\alpha(t)$ is clipped between zero and one.

A set of numerical simulations can be performed to compare this new routing strategy (4.19) with the user–equilibrium strategy (4.13). The network considered is represented in Fig. 4.12. The results of four scenarii with different inflows are represented in Table 4.10. The parameters of the controller were the same for the two strategies and are not particularly tuned in order to achieve the best TTS for this particular network. In all cases, the simulations last two hours. The length of road 1 is 5 km and the length of road 2 is 7 km. The function $V(\rho)$ used in the simulation is a simple linear relation such that $f^{\max} = 6000$ veh/h.

The four scenarii are:

1. At the two inputs, the inflows are constant (3000 veh/h). The two strategies lead to a constant splitting rate ($\alpha = 0.47$ for the system-optimum and $\alpha = 1$ for the user-optimum).

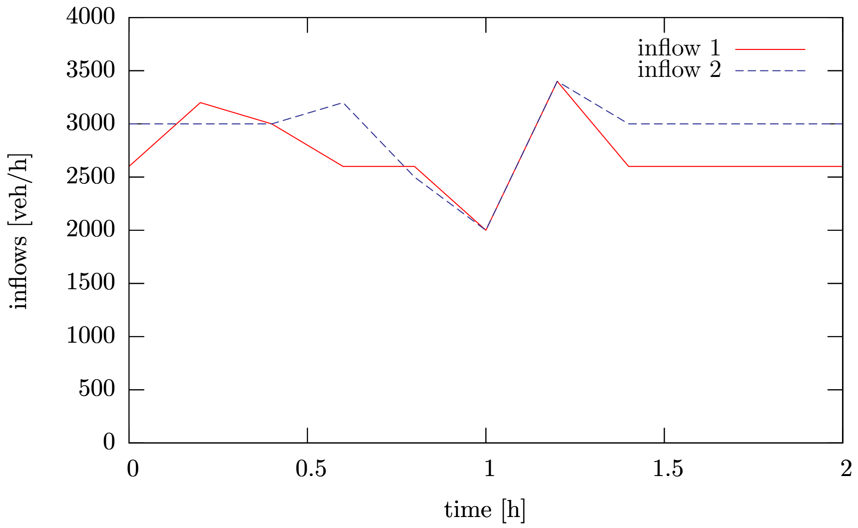


Figure 4.15: The inflows of the two inputs for scenarii 3 and 4.

2. At the two inputs, the inflows are constant (3200 veh/h for input 1 and 3000 veh/h for input 2). In this case, the two strategies produced two strongly different TTS. The reason is that the user-optimum strategy leads to a splitting rate of one: all drivers try to pass through road 1. But since the traffic on road 1 is limited to 6000 veh/h and 6200 veh/h are trying to pass, a traffic jam is growing before the bifurcation node.

3. The inflows of the two inputs vary with time (see Fig. 4.15). Sometimes the total inflow is higher than the capacity of road 1, sometimes it is lower.

4. The inflows are the same as in scenario 3 with the artificial appearance of a traffic jam to validate the use of the function g (4.17) when the density is higher than σ. This traffic jam appears by limiting the outflow of the output 1 to 300 veh/h during 6 minutes after 30 minutes of simulation.

As represented in Table 4.10, for the four scenarii, using the new routing strategy (4.19) allows a better TTS on the network at the end of the simulation.

4.4 Ramp metering with linear feedback

The second method developed in this work is a ramp metering strategy. This method is able to cope with non-steady state and consists in controlling the flow on an on-ramp. To limit temporarily the flow entering a highway can sometimes prevent the appearance of a traffic jam and thus increase the global efficiency. The ramp metering can also be used to prevent capacity drops at the junctions from which the system can be unable to recover (see [45]). This is generally performed using a traffic light. Controlling the red/green frequency allows to impose the desired inflow.

The idea developed in this section is to stabilise the system through linear feedback. Linear feedback means that the controlled inflows are linear function of measurements performed on the system. The stabilisation prevents the appearance of a traffic jam and is thus beneficial for the global efficiency of the network. For simplicity this strategy is only illustrated with the single order LWR model but could also be applied to the second order AR model (see Chapter 5).

4.4.1 Stability problem formulation

Consider a network with n roads and, initially, no congestion. This general network can contain loops and all the junctions described in Section 3.2. The absence of congestion ($\rho < \sigma$) implies that there is a unique correspondence between the density ρ and the flow $f(\rho)$. The single order model (3.4) can thus be rewritten in terms of the flows q_i:

$$\partial_t q_i(x,t) + c(q_i)\partial_x q_i(x,t) = 0 \qquad \forall i = 1, \ldots, n.$$

The absence of congestion also allows to have simpler expressions for the boundary conditions at the junction than the complete formulations of Section 3.2:

- For the merge of roads i and j to road k

$$q_k(0) = q_i(L_i) + q_j(L_j).$$

- For the divergence of road i to roads j and k

$$\begin{aligned} q_j(0) &= \alpha_i q_i(L_i) \\ q_k(0) &= (1-\alpha_i) q_i(L_i) \end{aligned}$$

  where α_i expresses the fraction of the drivers from road i wishing to take road j.

- For an origin junction at the beginning of road i

$$q_i(0) = \begin{cases} u_i & \text{for a control input through ramp metering} \\ v_i & \text{for an unknown disturbance.} \end{cases}$$

Since linear feedback is considered, the controlled entries are function of the traffic state. For example, if the input at the entrance of road i is controlled then the network operator can choose to make it dependent of the traffic state on road j

$$u_i = k_i q_j(L_j).$$

In this thesis it is assumed, without loss of generality, that the controls are function of the states at the end of some roads. It allows to apply the stability Theorem 2.3.2. If the state at the middle of a road must be used, it is always possible to cut it in two smaller roads connected by a 1-1 junction.

The complete system for the network can thus be expressed as

$$\frac{\partial}{\partial t}\begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix} + \begin{pmatrix} c(q_1) & & \\ & \ddots & \\ & & c(q_n) \end{pmatrix} \frac{\partial}{\partial x}\begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix} = 0 \qquad (4.20)$$

together with the boundary conditions

$$\begin{pmatrix} q_1(0,t) \\ \vdots \\ q_n(0,t) \end{pmatrix} = (A+K)\begin{pmatrix} q_1(L_1,t) \\ \vdots \\ q_n(L_2,t) \end{pmatrix} + V \qquad (4.21)$$

where

- A is a known matrix with entries $0, 1, \alpha_i$ or $(1-\alpha_i)$
- K is the control matrix which contains all the k_i parameters
- V includes the disturbance terms v_i

The values of the k_i are design parameters used by the operator to influence the system dynamics. Non-zero values can have positive effects on the system but too large values can produce instability. The main question is thus what are the upper bounds on the k_i such that the system remains stable.

If it is assumed that the uncontrolled perturbations v_i are constant, a stability condition for the system (4.20)–(4.21) can be obtained using Theorem 2.3.2 presented in Section 2.3. This theorem claims that if the initial state is close enough to an equilibrium and if

$$\rho(|A+K|) < 1, \qquad (4.22)$$

where ρ denotes the spectral radius, then the system is stable.

Condition (4.22) ensures local stability for the nonlinear system with constant perturbations. It can also be shown that this condition ensures stability for the linearised version of (4.20)–(4.21) with non-constant perturbations. Indeed, the transfer function $G(s)$ between the perturbations $v_i(t)$ and the outputs $q(1,t)$ can be evaluated as

$$G(s) = \left(I - (A+K)e^{-\Lambda s}\right)^{-1} e^{-\Lambda s} \qquad (4.23)$$

where

$$e^{-\Lambda s} = \begin{pmatrix} e^{-\frac{L_1}{v_1}s} & & 0 \\ & \ddots & \\ 0 & & e^{-\frac{L_n}{v_n}s} \end{pmatrix}$$

with v_i the speed of the drivers on road i. The system is stable if the input–output transfer function $G(s)$ has no pole in the right half-plane. This is the case if

$$\exists \epsilon > 0 \qquad s.a. \qquad \left|\det\left(I - (A+K)e^{-\Lambda s}\right)\right| \geq \epsilon \qquad \forall s \mid \mathrm{Re}(s) \geq 0. \quad (4.24)$$

Theorem 4.4.1. *If* (4.22) *holds then condition* (4.24) *is also satisfied.*

Proof. Since $\mathrm{Re}(s) \geq 0$,

$$\left|e^{-\frac{L_i}{v_i}s}\right| \leq 1$$

which implies

$$\left|(A+K)e^{-\Lambda s}\right| \leq |A+K|\,.$$

From [36, 5], it is known that $\rho(A) < \rho(|A|)$ and if $B \geq A \geq 0$ then $\rho(B) \geq \rho(A)$. This leads to

$$\rho\left((A+K)e^{-\Lambda s}\right) \leq \rho\left(\left|(A+K)e^{-\Lambda s}\right|\right) \leq \rho\left(|(A+K)|\right).$$

Since $\rho(|A+K|) = \mu < 1$ this implies that

$$\rho\left((A+K)e^{-\Lambda s}\right) \leq \mu < 1 \qquad \forall s \mid \mathrm{Re}(s) \geq 0.$$

The largest eigenvalue of $(A+K)e^{-\Lambda s}$ being strictly less than one on the whole right half-plane, there exists ϵ such that (4.24) holds. □

Condition (4.22) being a sufficient stability condition, it is interesting to describe the upper bound on K such that (4.22) is satisfied. From [5], it is known that

$$\rho(|A+K|) \leq \rho(|A| + |K|).$$

The stability is then assured if

$$\rho(|A| + |K|) < 1.$$

The question becomes "what is the upper bound on K such that $\rho(|A|+|K|) < 1$". This question is analysed in the next section and an application to a particular road network is then considered.

4.4.2 The spectral radius problem

Consider two positive matrices A and K where A is a known "nominal" part and K is function of some free parameters. The problem is to find the biggest set S such that

$$\rho(A+K) < 1 \qquad \forall K \in S.$$

One particular approach consists of considering structured matrices $K = E_1 \Delta E_2^T$ where Δ is the unknown disturbance and E_1 and E_2 are fixed matrices.

The problem is then to find the *stability radius of A with respect to nonnegative perturbations of structure (E_1, E_2^T)*; this radius is defined by

$$r_{\mathbb{R}_+}(A; E_1, E_2^T) = \inf\{\|\Delta\|\,; \Delta \geq 0, \rho(A + E_1 \Delta E_2^T) \geq 1\}.$$

All perturbations in the following set

$$S := \left\{E_1 \Delta E_2^T \,|\, \|\Delta\| < r_{\mathbb{R}_+}(A; E_1, E_2^T)\right\}$$

are then shown to yield a stable system $A + K$. This problem is solved in [34] and a computable formula is provided.

In this thesis, these results are extended into a particular direction. The perturbations matrices Δ considered are in the set $\mathcal{D}$ of nonnegative *diagonal* matrices $\mathcal{D} = \{\text{diag}\{k_1, \ldots, k_m\} \mid k_i \geq 0\}$. The parameters k_i are the so-called free parameters occurring in the matrix K, E_1 and E_2 are two matrices placing the elements in appropriate positions in K. The two matrices E_1 and E_2^T have the following properties: there is a non-zero element in row i and column j of E_1 if k_j is present in row i of K and of E_2^T if k_i is present in column j of K. This may be clarified by an example: if

$$K = \begin{pmatrix} 2k_1 & 0 & 0 \\ 0 & 0 & k_2 \\ k_1 & 0 & 0 \end{pmatrix},$$

then

$$\Delta = \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix}, \qquad E_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad E_2^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The matrices K are also restricted to the case where both E_1 and E_2 are nonnegative: $E_1 \geq 0,\ E_2 \geq 0$. Notice that if one of the parameters appears in several rows and columns, it will be repeated several times in the diagonal matrix Δ.

The problem is to find the biggest set $S_{\mathcal{D}} \subseteq \left\{E_1 \Delta E_2^T \;\middle|\; \Delta \in \mathcal{D}\right\}$ containing the origin and all the perturbations such that the system remains stable:

$$S_{\mathcal{D}} := \left\{K = E_1 \Delta E_2^T \mid \Delta \in \mathcal{D},\ \rho(A + K) < 1\right\}$$

where A, E_1, E_2, Δ are nonnegative matrices. It can be pointed out that this is in fact a starlike set.

Theorem 4.4.2. *The set $S_{\mathcal{D}}$ is a starlike set.*

Proof. From [5], it is known that if $A, B \geq 0$ then $\rho(A+B) \geq \rho(A)$. It implies that if $K \in S_{\mathcal{D}}$

$$\rho(A + K) < 1 \Rightarrow \rho(A + \alpha K) < 1 \qquad \forall \alpha \quad 0 \leq \alpha \leq 1$$

and $\alpha K \in S_{\mathcal{D}}$. If $K \notin S_{\mathcal{D}}$ then

$$\rho(A + K) \geq 1 \Rightarrow \rho(A + \beta K) \geq 1 \qquad \forall \beta \quad 1 \leq \beta$$

and $\beta K \notin S_{\mathcal{D}}$. $\square$

Since the spectral radius is a continuous function of the parameter K, the boundary of the set $S_{\mathcal{D}}$ is described by

$$\partial S_{\mathcal{D}} = \{K \mid \exists i,\ k_i = 0 \ \&\ \rho(A+K) \leq 1 \qquad \text{or} \qquad \rho(A+K) = 1 \ \&\ K \geq 0\}$$

Examples show later that this set is in general not convex.

On the one side, the problem solved in [34] is more general because it does not assume that the perturbation Δ is diagonal. But, on the other side, when Δ is diagonal, their problem is more restrictive than the one addressed in this work. All the operator norms induced by an arbitrary monotonic norm on $\mathbb{R}^n$ of a diagonal matrix are equal to the maximum of the elements of the matrix (see [36]). It means that the set S considered in [34] is a box with $k_i \leq k_i^m$ and $k_1^m = \ldots = k_n^m$ and hence a convex subset of $S_{\mathcal{D}}$. In this work, it is shown that there exists a maximum starlike set $S_{\mathcal{D}} = \{E_1 \Delta E_2^T \mid \Delta \in \mathcal{D}\}$ for which all matrices K in $S_{\mathcal{D}}$ lead to stable $A+K$ and the boundary of this set is described in terms of polynomial equations.

First some new theoretical results are developed and the corresponding algorithm to compute $\partial S_{\mathcal{D}}$ is then presented.

Theoretical results

In this section it is shown that the problem may be decoupled in smaller subproblems involving only a subset of the parameters k_i. To each of these subproblems corresponds a starlike set $S_{\mathcal{D}_i}$ for which an analytical expression can be obtained. The set $S_{\mathcal{D}}$ is the intersection of the sets $S_{\mathcal{D}_i}$.

Since $K = E_1 \Delta E_2^T$ is nonnegative and since the eigenvalues are continuous functions of the matrix elements, one has that the critical switch between $\rho < 1$ and $\rho \geq 1$ occurs when

$$\rho(A + E_1 \Delta E_2^T) = 1.$$

Working only with positive matrices, the spectral radius is also an eigenvalue and hence the above condition is equivalent to

$$\det(A + E_1 \Delta E_2^T - I) = 0$$

and

$$\det(E_1 \Delta E_2^T - (I - A)) = 0.$$

Since $\det(I - A) \neq 0$ ($\rho(A) < 1$) the previous equation can by multiplied by $\det(I - A)^{-1}$ to obtain

$$\det((I - A)^{-1} E_1 \Delta E_2^T - I) = 0.$$

Using the fact that

$$\det(MN - I) = 0 \Leftrightarrow \det(NM - I) = 0$$

this is also equivalent to

$$\det(E_2^T (I - A)^{-1} E_1 \Delta - I) = 0$$

where $M := E_2^T(I-A)^{-1}E_1$ is nonnegative since $(I-A)^{-1} = \sum_{i=1}^{\infty} A^i$ and E_1, E_2 are nonnegative. The Lemma 4.4.3 (see [17]) can be used to put M under a normal form $\hat{M}$.

Lemma 4.4.3. *Every nonnegative matrix A has a normal form which can obtained under congruent permutations:*

$$\hat{A} = PAP^T = \begin{pmatrix} \hat{A}_{11} & & 0 \\ & \ddots & \\ * & & \hat{A}_{mm} \end{pmatrix} \tag{4.25}$$

where each diagonal block $\hat{A}_{ii}$ is square, irreducible or just a 1×1 zero block.

Applying the same permutation to Δ, one can define $\hat{\Delta} = P\Delta P^T$ and have

$$\det(\hat{M}\hat{\Delta} - I) = 0.$$

This clearly decomposes in a number of decoupled problems

$$\det(\hat{M}_{ii}\hat{\Delta}_{ii} - I) = 0.$$

This polynomial equation describes a part of $\partial S_{\mathcal{D}_i}$. The intersection of all these sets leads to the admissible set $S_{\mathcal{D}}$. The different subproblems can be considered:

- Let $\hat{M}_{ii} = 0$ then $\det(\hat{M}_{ii}\hat{\Delta}_{ii} - I) \neq 0$ for all *bounded* $\hat{\Delta}_{ii}$;
- Let $\hat{M}_{ii} \neq 0$ and irreducible. If the size n_i of $\hat{M}_{ii} = [m_{r,c}]_{r,c=1}^{n_i}$ is small enough, the problem can be exactly solved.
  - If $n_i = 1$, the solution is trivial $\det(m_{11}k_1 - 1) = 0 \quad \text{for } k_1 = m_{11}^{-1}$.
  - If $n_i = 2$ then $\det\left(\begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22}\end{pmatrix}\begin{pmatrix} k_1 & 0 \\ 0 & k_2\end{pmatrix} - I\right) = 0$ or equivalently:

    $$(m_{11}m_{22} - m_{12}m_{21})k_1k_2 - m_{11}k_1 - m_{22}k_2 + 1 = 0. \tag{4.26}$$

    The stable region for the k_1, k_2 is thus a starlike set with boundary defined by (4.26) and $k_{1,2} = 0$ (see Fig. 4.16).
  - If $n_i = 3$, one has

    $$\det(\hat{M}_{ii}\hat{\Delta}_{ii} - I) = 0$$

    or equivalently:

    $$\det(\hat{M}_{ii})k_1k_2k_3 + (m_{21}m_{12} - m_{11}m_{22})k_1k_2 + (m_{13}m_{31} - m_{11}m_{33}) \\ k_1k_3 + (m_{23}m_{32} - m_{22}m_{33})k_2k_3 + m_{11}k_1 + m_{22}k_2 + m_{33}k_3 - 1 = 0. \tag{4.27}$$

    The stable region for the k_i is thus also a starlike set whose boundaries are defined by (4.27) and $k_{1,2,3} = 0$ (see Fig. 4.17).

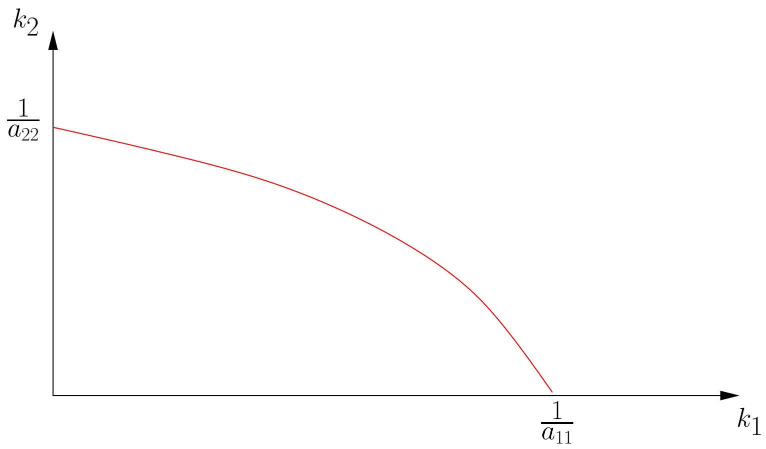


Figure 4.16: The largest set of the parameters k_1 and k_2 containing the origin such that $A + E_1\Delta E_2^T$ is stable.

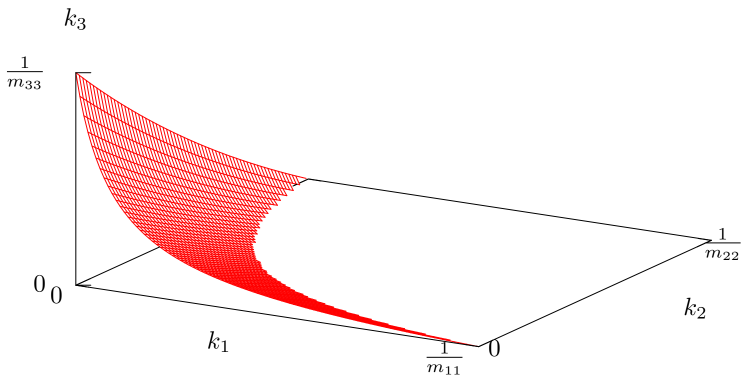


Figure 4.17: The boundary of the largest set (k_1, k_1, k_3) containing the origin such that $A + E_1\Delta E_2^T$ is stable

– It may happen that a coefficient k_i appears in different blocks $\hat{\Delta}_{ii}$. For example, if

$$A = \begin{pmatrix} a_{11} & a_{12} & & \\ a_{21} & a_{22} & & \\ & & a_{33} & a_{34} \\ & & a_{43} & a_{44} \end{pmatrix}, \quad K = \begin{pmatrix} k_1 & & & \\ & k_2 & & \\ & & k_3 & \\ & & & k_1 \end{pmatrix}$$

then

$$\hat{\Delta}_{11} = \begin{pmatrix} k_1 & \\ & k_2 \end{pmatrix}, \quad \hat{\Delta}_{22} = \begin{pmatrix} k_3 & \\ & k_1 \end{pmatrix}.$$

In this case, the admissible set for (k_1, k_2, k_3) is simply the intersection of the two sets obtained by analysing the two subproblems. This is illustrated in Figure 4.18.

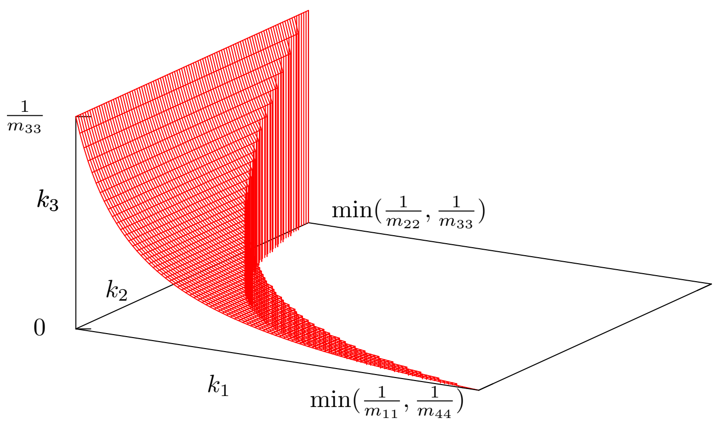


Figure 4.18: The admissible set is the intersection of the two admissible sets.

Modified problem

In the above analysis the set $S_{\mathcal{D}_i}$ can become quite complex to describe if n_i becomes large, since one has to solve a polynomial equation of degree n_i in several variables. Two approximations of the original problem that are easier to compute are described here. The approximation consists in finding necessary conditions to ensure stability (i.e. a subset of $S_{\mathcal{D}_i}$).

For example, one can freeze one particular k_i and express the conditions on the remaining parameters. This subset will be a slice of $S_{\mathcal{D}_i}$. This subset is still starlike in the leftover parameters.

One other possibility is to express a condition on the maximum of the k_i in the same way as [34]. The following result is a refinement of the global bound obtained in [34]. Let ρ_i be the spectral radius of M_{ii} then

$$\begin{aligned} \det(\hat{M}_{ii}\hat{\Delta}_{ii} - I) &\neq 0 \qquad \text{for } \hat{\Delta}_{ii} < \rho_i^{-1} I \\ \det(\hat{M}_{ii}\hat{\Delta}_{ii} - I) &= 0 \qquad \text{for } \hat{\Delta}_{ii} = \rho_i^{-1} I. \end{aligned}$$

Proof. Let x_i be the Perron vector of the irreducible matrix $\hat{M}_{ii}$. It is well known (see [17]) that for an irreducible matrix the so-called Perron vector x_i corresponding to the (positive) Perron root ρ_i is strictly positive. Therefore

$$\hat{M}_{ii} x_i = \rho_i x_i, \;\; x_i > 0$$

then clearly

$$(\hat{M}_{ii}\rho_i^{-1} I - I)x_i = 0 \qquad \hat{\Delta}_{ii} = \rho_i^{-1} I.$$

Also for $\hat{\Delta}_{ii} < \rho_i^{-1} I$

$$\det(\hat{M}_{ii}\hat{\Delta}_{ii} - I) \neq 0$$

since there exists a scaling

$$\left\| D^{-1}\hat{M}_{ii} D \right\|_\infty = \rho_i$$

and clearly

$$\left\| D^{-1}\hat{M}_{ii}\hat{\Delta}_{ii} D \right\|_\infty = \left\| D^{-1}\hat{M}_{ii} D \hat{\Delta}_{ii} \right\|_\infty < 1.$$

□

Therefore it can be claimed that all matrices Δ in the following set

$$S = \left\{ \Delta \mid \begin{pmatrix} \hat{\Delta}_{11} & & \\ & \ddots & \\ & & \hat{\Delta}_{mm} \end{pmatrix} = P\Delta P^T, \right.$$
$$\left. \hat{\Delta}_{ii} < \begin{cases} \text{any bounded value} & \text{if } \rho_i = 0 \\ \rho_i^{-1} I & \text{if } \rho_i \neq 0 \end{cases} \right\}$$

are such that

$$\rho(A + E_1 \Delta E_2^T) < 1.$$

The problem may thus be split into several subproblems. If the subproblems are small enough, one has some analytical necessary and sufficient conditions. If the subproblems are more complex, to ensure that $\rho(A + E_1\Delta E_2^T) < 1$, sufficient conditions may be used such as freezing a k_i or imposing that for each $\hat{M}_{ii} \neq 0$, $\hat{\Delta}_{ii} < \rho_i^{-1} I$.

Algorithm

The results presented in the previous section can be used to construct the set $S_{\mathcal{D}}$, in the following manner:

1. Compute the matrix $M := E_2^T(I-A)^{-1}E_1$ and perform permutations to put it under the normal form (4.25). This can be done by applying the following algorithm :
   (a) Use Tarjan's algorithm [58] to find the set of strongly connected subgraphs associated to the graph G defined by the Adjacency Matrix M^{ad} ($M^{ad}_{i,j}=1$ if $M_{i,j}\neq 0$, $M^{ad}_{i,j}=0$ otherwise).
   (b) Consider a new graph G' whose nodes represent the strongly connected subgraphs: two nodes i and j of G' are connected if there exists one edge between a node of G in the subgraph i and a node of G in the subgraph j. The adjacency matrix of this new graph G' can be computed simply from M^{ad} by first summing up the rows corresponding to the same subgraph and then summing up the columns corresponding to the same subgraph.
   (c) Identify a leaf i of the graph G' (which always exists because there is no cycle in G') and permute the columns and the rows of M corresponding to the subgraph i at the beginning of the matrix. Suppress node i from G'. Repeat 1c until M is in canonical form (4.25).
2. For each of the $\hat{M}_{ii}$ blocks, express the condition $\det(\hat{M}_{ii}\hat{\Delta}_{ii}-I)=0$ which describes a part of the boundary of $S_{\mathcal{D}}$. If the size of $\hat{M}_{ii}$ is too high, a more restrictive condition such as first freezing a k_i or a condition in term of $\rho(\hat{M}_{ii})$ can be used.

It can be now claimed that, if

$$(k_1,\ldots,k_n)\in S_{\mathcal{D}}$$

then

$$\rho\left(A+E_1\begin{pmatrix}k_1 & & \\ & \ddots & \\ & & k_n\end{pmatrix}E_2^T\right)<1.$$

4.4.3 A numerical example

Consider the network represented in Figure 4.19. Assuming that there is no congestion, the traffic evolution is modelled by

$$\frac{\partial}{\partial t}\begin{pmatrix}q_1\\ \vdots\\ q_9\end{pmatrix}+\begin{pmatrix}c(q_1) & & \\ & \ddots & \\ & & c(q_9)\end{pmatrix}\frac{\partial}{\partial x}\begin{pmatrix}q_1\\ \vdots\\ q_9\end{pmatrix}=0$$

together with the boundary conditions

$$\begin{aligned} q_1(0) &= v_1 \\ q_2(0) &= \alpha q_1(L_1) \\ q_3(0) &= (1-\beta) q_2(L_2) \\ q_4(0) &= q_3(L_3) + u_2 \\ q_5(0) &= (1-\gamma) q_4(L_4) \\ q_6(0) &= q_5(L_5) + u_3 \\ q_7(0) &= (1-\alpha) q_1(L_1) \\ q_8(0) &= q_7(L_7) + u_4 \\ q_9(0) &= q_6(L_6) + q_8(L_8) \end{aligned}$$

where α, β, γ are the splitting factors at the three diverging junctions ($\alpha = 0.8,\ \beta = 0.1,\ \gamma = 0.2$).

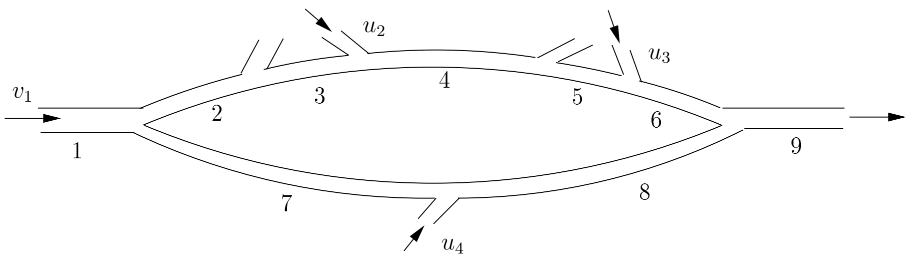


Figure 4.19: A simple network where three on-ramps are controlled by ramp-metering.

If the network operator chooses to control the inflows using the following linear feedback

$$\begin{aligned} u_2 &= k_2 q_6(L_6) \\ u_3 &= k_3 q_6(L_6) \\ u_4 &= k_4 q_8(L_8) \end{aligned}$$

then the boundary conditions can be rewritten under the following form

$$\begin{pmatrix} q_1(0) \\ q_2(0) \\ q_3(0) \\ q_4(0) \\ q_5(0) \\ q_6(0) \\ q_7(0) \\ q_8(0) \\ q_9(0) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & (1-\beta) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & k_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & (1-\gamma) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & k_3 & 0 & 0 & 0 \\ (1-\alpha) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & k_4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} q_1(L_1) \\ q_2(L_2) \\ q_3(L_3) \\ q_4(L_4) \\ q_5(L_5) \\ q_6(L_6) \\ q_7(L_7) \\ q_8(L_8) \\ q_9(L_9) \end{pmatrix}. \tag{4.28}$$

The system will be stable if the spectral radius of the matrix in (4.28) is less than one. The bounds on the k_i values can be obtained using the results presented in Section 4.4.2. In this particular case, the matrix $\hat{M}$ introduced in the previous section is

$$\hat{M} = \begin{pmatrix} 0.8000 & 1.0000 & 0 \\ 0.8000 & 1.0000 & 0 \\ 0 & 0 & 1.0000 \end{pmatrix}.$$

It can be seen that the problem can be decomposed in two subproblems with a bound on k_2, k_3 and another bound on k_4

$$\begin{aligned} 0.8\,|k_2| + |k_3| &< 1 \\ |k_4| &< 1. \end{aligned}$$

This set of value for k_i which ensures stability is represented in Figure 4.20.

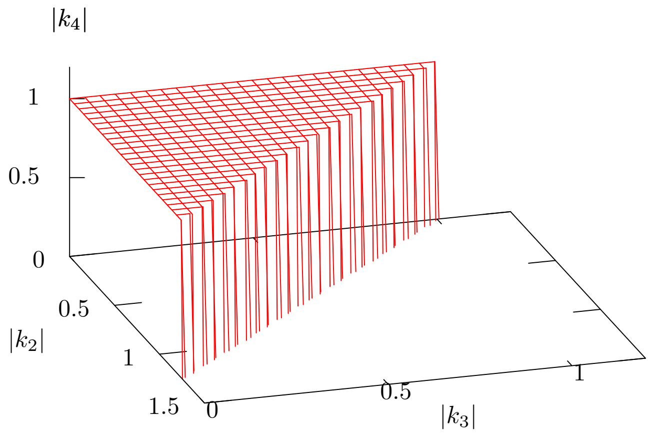


Figure 4.20: The boundary of the set $S_{\mathcal{D}}$ such that for all k_i in this set the system is stable.

4.4.4 Selection of the parameters

The selection of the appropriate values for the free parameters k_i is a difficult problem. A possible criterion is to minimise the probability of a traffic jam appearance at the end of the network. This probability is linked to increase of the network outflow produced by the uncontrolled perturbations v_i.

A usual way consists in linearising the system under consideration and minimising a norm of the input–output transfer function (4.23) which links an uncontrolled perturbation v_i to the output. Two particular norms are:

2-norm

$$\|G\|_2 := \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} |G(j\omega)|^2 \, d\omega\right)^{1/2} \tag{4.29}$$

∞-norm

$$\|G\|_\infty := \sup_{\mathrm{Re}(s)>0} |G(s)| \, . \tag{4.30}$$

These norms are used since they allow to obtain bounds on the output in function of the input (see [15]):

If

$$y(s) = G(s)v(s)$$

then

$$\begin{aligned} \|y\|_\infty &\leq \|G\|_2 \, \|v\|_2 \\ \|y\|_2 &\leq \|G\|_\infty \, \|v\|_2 \, . \end{aligned}$$

The minimisation of the 2-norm does not allow to select the k_i parameters. Indeed the variable s only appear in $G(s)$ through exponentials, the integral (4.29) is thus not converging and the 2-norm is always infinite.

The minimisation of the ∞-norm provides paradoxical results. For example consider the network represented in Fig. 4.21 with the control

$$u(t) = -ky(t).$$

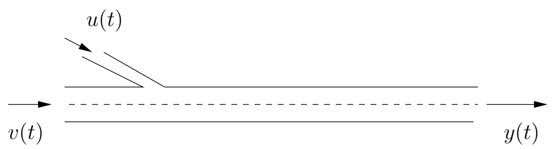


Figure 4.21: A highway with a controlled on-ramp.

One can expect that a value of k around 0.5 leads to a better response than without control. The corresponding transfer function is

$$G(s) = \frac{e^{-s(\tau_1+\tau_2)}}{1 + ke^{-s(\tau_2+\tau_3)}}$$

where τ_i is a time delay. Since

$$|G(j\omega)| = \frac{1}{\sqrt{(1+k^2) + 2k\cos(\omega(\tau_1+\tau_2))}}$$

the ∞-norm (4.30) is minimised for $k = 0$. It indicates that the best solution is no control at all! This surprising result can be explained by observing that for $k \neq 0$ there exists a very particular input v such that the output is worse than without control. Since minimising the ∞-norm corresponds to minimising the worst case, it leads to no control. For the network represented in Fig. 4.21 there exists a particular perturbation represented in Fig. 4.22 which produces an output worst (in term of the 2-norm) than without control.

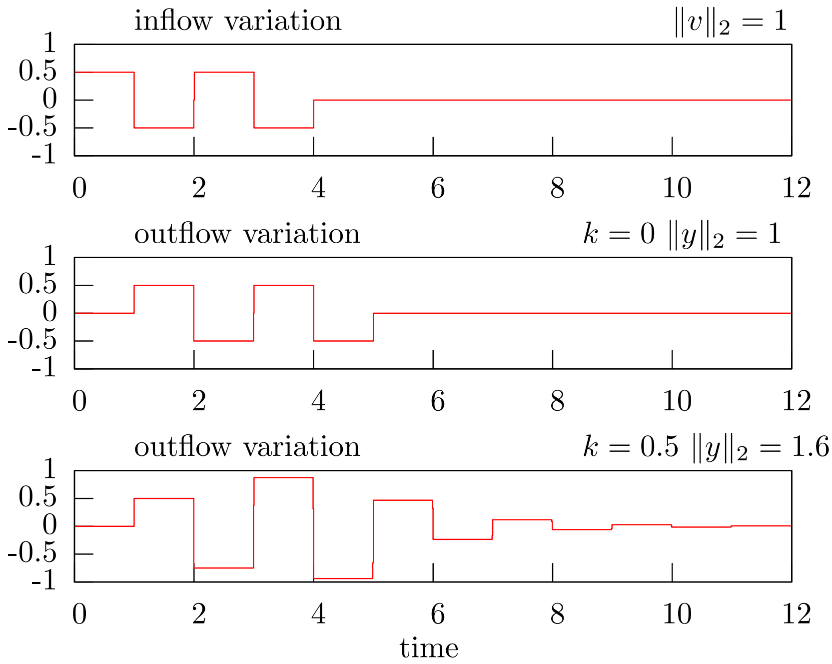


Figure 4.22: The response of the system to a particular input with ($k = 0.5$) and without ($k = 0$) control.

Even for some more complex networks such as the network represented in Fig. 4.19, the minimisation of the ∞-norm leads to $k_2 = k_3 = k_4 = 0$.

A criterion based on the worst-case minimisation is not the most significant criterion for the considered traffic problems. If it is assumed that the shape of the perturbation occurring at the network entrance is known, then it is possible to select values for the free parameters k_i which optimise the system behaviour. This assumption is not purely theoretical since some perturbations can reasonably be foreseen. A great part of the commuting traffic is daily recurrent. One can think of the peak hours near big cities.

For example, one can consider the network represented in Fig. 4.21 and the perturbation in Fig. 4.23. This corresponds to a situation where the ramp–metering is used to handle a temporary increase of flow on the highway.

Keeping as main objective to minimise the probability of a traffic jam ap-

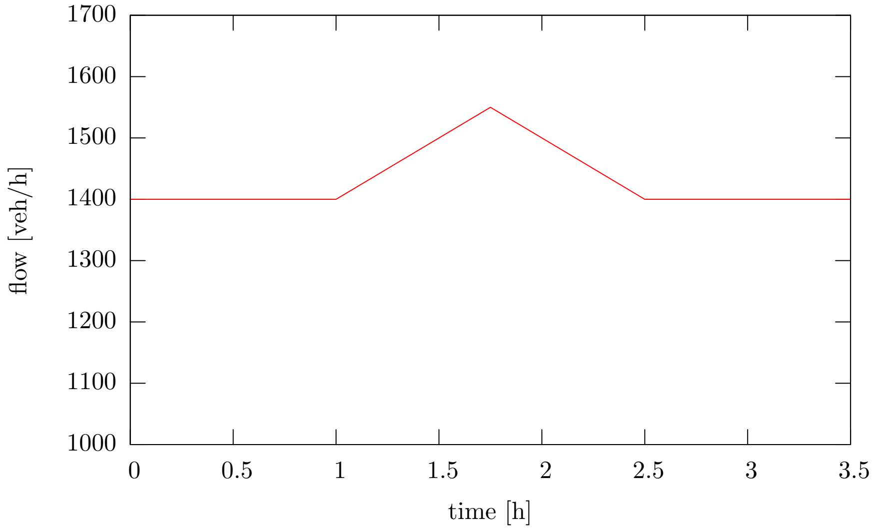


Figure 4.23: The flow at the entrance of the highway.

pearance at the end of the network, one can use the following cost function

$$J(T) = \int_0^T \left(\alpha_1 \max(y(t), 0)^2 + \alpha_2 n(t)\right) dt$$

where $n(t)$ is the number of vehicle waiting at the controlled on-ramp. The risk of a traffic jam is linked to the increase of the outflow $(\max(y(t), 0))$ and the large increase are more risky and therefore more penalised $(\max(y(t), 0)^2)$. The cost of the users waiting at the on-ramp (proportional to $\int n(t)dt$) must also be taken into account. The two coefficients α_1 and α_2 can be used to balance the risk and the cost.

For $\alpha_1 = \alpha_2 = 1$; the value of k minimising $J(T)$ is 0.92 and the corresponding output is represented in Fig. 4.24. It can be noticed that the maximal amplitude of the perturbation is reduced with control. The network operator can also choose to lessen the oscillations by reducing the coefficient k but it will be at the cost of an increase of the traffic jam probability.

Figure 4.24: The response of the system to a common input with ($k = 0.47$) and without ($k = 0$) control.

Chapter 5

A second order network traffic model

The LWR first order model is already able to represent various phenomena but not all of them. The assumption of an algebraic relation between the density and the speed implies that the drivers face infinite accelerations when crossing density shocks and that they don't anticipate what happens in front of them. This is clearly unrealistic. Moreover the LWR model is unable to predict the emergence of phantom traffic jams or stop-and-go waves. The addition of the g function in the expression of the Receiving capacity (3.10) to represent the capacity drop phenomenon is more an ad-hoc correction than a natural extension.

To overtake these limitations new models with more complexity must be developed. The idea is to add a second PDE to describe the evolution of the speed. Various second order models have been developed (see [54, 60, 40]) but most of them have important limitations such as the appearance of negative speed (see [10]). The model proposed and extended in this work is the model of Aw & Rascle (AR) which does not contain such limitation (see [2]).

In Section 5.1 the AR single road model is introduced and some of its main properties are presented. A stability analysis is then performed in Section 5.2.

To obtain a model for a road network one still have, like in Chapter 3, to describe what happens at the junctions. This description is done, like for the single order model, by the resolution of a Riemann problem at the junction. The solution of this Riemann problem is obtained in two steps:

1. A description of the admissible regions.

2. The introduction of additional conditions to select appropriate values in the admissible regions.

The core of a junction model is the additional conditions which are introduced in the second step. Each new assumption leads to a different model with particular properties and complexity. Different models have already been proposed in the literature such as [32, 33] and [20]. It turns out that the model proposed in this thesis [23] is very close to another model [50] which has been independently and simultaneously developed even if the formulations appear quite different. Once the formulation of the new junction is done the properties of our model will be analysed. Finally this junction model is compared with other existing models.

5.1 Second order model for an infinite road

For a second order model, the evolution of the traffic state is described by two equations. The first one is, like for the LWR model, derived from the flow conservation assumption:

$$\partial_t \rho + \partial_x(\rho v) = 0. \tag{5.1}$$

The second one describes the evolution of the speed of the drivers. The idea is that the variation of the speed is linked to the variation of the density through a function $p(\rho)$:

$$(\partial_t + v\partial_x)v + (\partial_t + v\partial_x)p(\rho) = 0. \tag{5.2}$$

The function $p(\rho)$ must be a smooth increasing function in order to reduce the speed when the density increases. The use of the Lagrangian derivative $(\partial_t + v\partial_x)$ instead of the simple time derivative ∂_t allows to take into account the variation of density really perceived by the drivers. It is usually assumed that

$$\frac{d^2}{d\rho^2}(\rho p(\rho)) > 0. \tag{5.3}$$

For example this condition is satisfied if $p(\rho) = v^{\max} - V(\rho)$. In this case the AR model coincides with Zhang's model proposed in [61]. From now on it is assumed that (5.3) holds.

One additional term may be added to the right hand side of (5.2), the relaxation term $\frac{V(\rho)-v}{\tau}$. This term expresses the fact that the drivers tend to accelerate or break to reach a preferential speed V function of the surrounding density ρ. The dynamics induced by this term are usually slow and are negligible when dealing with very short time such as the crossing of discontinuities or a junction. This relaxation term is therefore not used when the Riemann problem on a single road or at a junction is analysed. However its influence on the stability will be analysed in Section 5.2.

Multiplying (5.1) by $(v+p(\rho))$ and (5.2) by ρ and adding these two equations leads to

$$\partial_t(\rho(v + p(\rho))) + \partial_x(\rho v(v + p(\rho))) = 0.$$

Therefore the system is composed of two conserved quantities: ρ and $\rho(v + p(\rho))$. The AR model corresponds thus to the system (2.13)–(2.14), used in

Section 2.4.4 to illustrate the different waves, with $u_1 = \rho,\ u_2 = v$ and $y = \rho(v + p(\rho))$. The corresponding Jacobian matrix is

$$A = \begin{pmatrix} v & \rho \\ 0 & v - \rho p'(\rho) \end{pmatrix}$$

and the two associated eigenvalues are

$$\begin{aligned} \lambda_1 &= v - \rho p'(\rho) \\ \lambda_2 &= v. \end{aligned}$$

Since $\lambda_1 \leq \lambda_2 \leq v$, according to Remark 2.2.1, it implies that the drivers do not react in function of what happen behind them.

As seen in Section 2.4.4, the solution of a Riemann problem for the AR system

$$(\rho(x,0), v(x,0)) = \begin{cases} U_l = (\rho_l, v_l) & \text{if } x < 0 \\ U_r = (\rho_r, v_r) & \text{if } x \geq 0 \end{cases}$$

consists of the connection of one rarefaction or shock wave with a contact-discontinuity. The solutions of the two most simple and common Riemann problems are illustrated in Fig. 5.1. For the complete description of the Riemann problem see [2] and Section 2.4.4. In the first two graphs of Figure 5.1, are represented in the $(\rho, \rho v)$ plane:

- the two initial state U_l and U_r;
- the straight line $(\rho, \rho v_r)$ passing through 0 and U_r;
- the curve $\Upsilon = (\rho, \rho K - \rho p(\rho))$ passing through U_l. All the points on this curve have the same value $v + p(\rho)$. There is only one value of K such that this curve passes through U_l, this value is equal to $v_l + p(\rho_l)$;
- the intermediate state U_c which is at the intersection of the straight line and Υ.

Two cases must be considered:

- If $\rho_l < \rho_c$, then the solution consists of a shock wave connecting U_l to U_c followed by a contact discontinuity connecting U_c to U_r. See Figure 5.1(a). The first graph represents the three states in the $(\rho, \rho v)$ plane, the second graph the initial state in the (x, ρ) plane and the third graph the state on the road after some time. The speed of the shock wave is equal to the slope of the line connecting U_l to U_c.
- If $\rho_l > \rho_c$, then the solution consists of a rarefaction wave connecting U_l to U_c followed by a contact discontinuity connecting U_c to U_r. See Figure 5.1(b). The space interval occupied by the rarefaction wave is

$$\left[\left. \frac{d(\rho K - \rho p(\rho))}{d\rho} \right|_{\rho=\rho_l} t,\ \left. \frac{d(\rho K - \rho p(\rho))}{d\rho} \right|_{\rho=\rho_c} t \right]$$

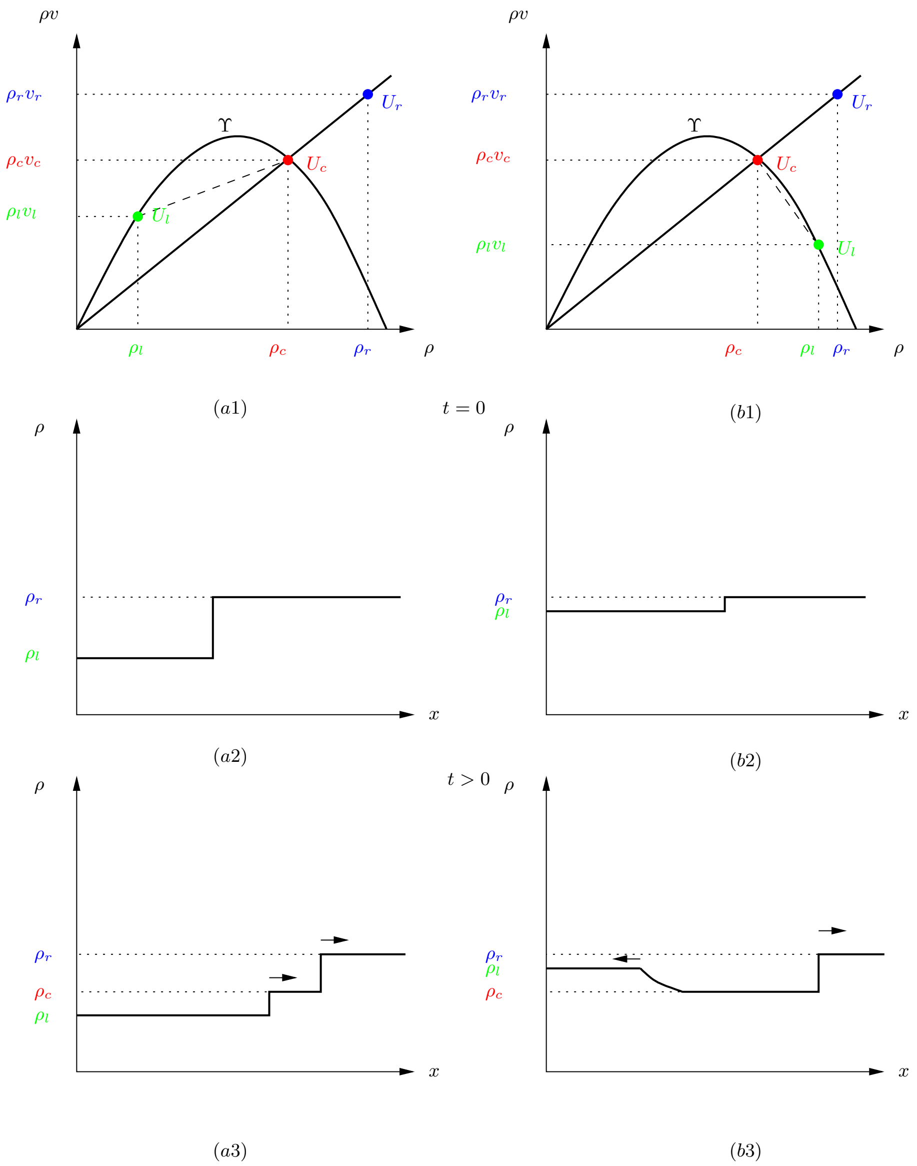


Figure 5.1: Two cases are represented in this figure. The first case (a) brings into play a shock wave and a contact discontinuity. In the second case (b) a rarefaction wave and a contact discontinuity are present. For each case three graphs are represented. In the first row, the two initial states U_l, U_r and the intermediate state U_c are represented. The intermediate state is at the intersection of the curve Υ and the straight line passing trough 0 and U_r. The second row represents the corresponding initial states in the (x, ρ) plane. The states of the densities after some time are represented in the last row.

which corresponds to t times the slope of the curve Υ evaluated at U_l and U_c.

For some initial values U_l, U_r, the solution can imply the vacuum ($\rho = 0$). In this case the system is no more strictly hyperbolic ($\lambda_1 = v - \rho p'(\rho) = v = \lambda_2$) and there is not unicity of the solution. However the differences may only appear on the speed when the density is equal to zero. In that case, from an engineering point of view, the speed is neither defined nor important since all the costs (4.5) are balanced with the density. All the solutions are therefore equivalent if the focus is on the cost or the "efficiency" aspect.

It can be shown that the single infinite road model (5.1)–(5.2) admits an invariant region

$$\mathcal{R} := \{(\rho, v) \mid 0 \leq v + p(\rho) \leq v_m, \quad \rho \geq 0,\ 0 \leq v \leq v_m\} \tag{5.4}$$

which ensures that the speed remains positive and that there exists a maximal speed and density.

5.1.1 Piecewise LWR

Assume that the quantity $v + p(\rho)$ is initially piecewise constant. Since this quantity is conserved along vehicle trajectory and that the only waves carrying a change of this quantity are contact-discontinuities, it turns out that $v + p(\rho)$ remains piecewise constant for $t > 0$ (see [41]).

On the domain of the (x, t) plane where $v + p(\rho) = K_i$, (5.1) leads to

$$\partial_t \rho + \partial_x(\rho(K_i - p(\rho))) = 0.$$

This equation expresses the fact that the traffic is described on each domain by the LWR model with the particular flow function

$$f(\rho) = -\rho p(\rho) + \rho K_i. \tag{5.5}$$

If the particular function $p(\rho) = -V(\rho)$ is used then the analogy is stronger since (5.5) becomes

$$f(\rho) = \rho V(\rho) + \rho K_i.$$

It means that this function f is the fundamental diagram shifted proportionally to $K_i = v - V(\rho)$ which represents the difference between the drivers' speed and the preferential speed.

5.2 Stability analysis

As already mentioned, a relaxation term $\frac{V(\rho)-v}{\tau}$ may be added to represent the fact that the driver want to reach a preferential speed function of the surrounding density. In that case, the system becomes

$$\begin{aligned} \partial_t \rho + \partial_x(\rho v) &= 0 \\ (\partial_t + v\partial_x)v + (\partial_t + v\partial_x)p(\rho) &= \frac{V(\rho)) - v}{\tau}. \end{aligned}$$

The introduction of this relaxation term can have an important influence on the stability of a homogeneous equilibrium $(\rho_e, v_e = V(\rho_e))$. The effect of a relaxation term on the stability of a general strictly hyperbolic system can easily be analysed.

Consider the following system

$$U_t + A(U)U_x = s(U) \tag{5.6}$$

where

$$U = \begin{pmatrix} \rho \\ v \end{pmatrix} \quad A(U) = \begin{pmatrix} a_{11}(U) & a_{12}(U) \\ a_{21}(U) & a_{22}(U) \end{pmatrix}, \quad s(U) = \begin{pmatrix} 0 \\ \frac{V(\rho)-v}{\tau} \end{pmatrix}.$$

The influence of a small perturbation (ζ, η) can be analysed by injecting $(\rho_e + \zeta, v_e + \eta)$ in (5.6) and performing a first order development:

$$\begin{pmatrix} \zeta \\ \eta \end{pmatrix}_t + A(U_e)\begin{pmatrix} \zeta \\ \eta \end{pmatrix}_x = \begin{pmatrix} 0 \\ \frac{V'(\rho_e)\zeta - \eta}{\tau} \end{pmatrix}. \tag{5.7}$$

From the first equation of (5.7), it can be deduced that

$$\eta_x = -\frac{\zeta_t + a_{11}\zeta_x}{a_{12}} \tag{5.8}$$

where a_{ij} denotes $a_{ij}(U_e)$. Taking the spatial derivative of the second equation of (5.7) leads to

$$\eta_{tx} + a_{21}\zeta_{xx} + a_{22}\eta_{xx} = \frac{V'(\rho_e)\zeta_x - \eta_x}{\tau}.$$

This equation can be rewritten replacing η_x using (5.8):

$$-\frac{\zeta_{tt} + a_{11}\zeta_{xt}}{a_{12}} + a_{21}\zeta_{xx} - a_{22}\frac{\zeta_{tx} + a_{11}\zeta_{xx}}{a_{12}} = \frac{V'(\rho_e)\zeta_x + \frac{\zeta_t + a_{11}\zeta_x}{a_{12}}}{\tau}$$

which is equivalent to

$$(\partial_t + (a_{11} + a_{12}V'(\rho_e))\partial_x)\zeta = -\tau(\zeta_{tt} + (a_{11} + a_{22})\zeta_{xt} + (a_{11}a_{22} - a_{12}a_{21})\zeta_{xx}).$$

Denoting by λ_1, λ_2 the eigenvalues of the matrix $A(U_e)$ and $c = a_{11} + a_{12}V'(\rho_e)$, the previous equation can be rewritten as

$$(\partial_t + c\partial_x)\zeta = -\tau\left((\partial_t + \lambda_1\partial_x)(\partial_t + \lambda_2\partial_x)\right)\zeta \tag{5.9}$$

describing the evolution of the perturbation.

It is natural to assume that the perturbation is local and bounded which ensures that the Fourier transform of ζ exists

$$\zeta(x,t) = \int_{-\infty}^{\infty} F(k,t) e^{ikx} dk.$$

Since ζ satisfies a linear equation (5.9), the $F(k,t)$ term can be expressed as

$$F(k,t) = Ae^{-i\omega t}$$

where the values ω and A are linked to the values of k. The perturbation ζ can thus be written as

$$\zeta(x,t) = \int_{-\infty}^{\infty} A(k) e^{ikx - i\omega(k)t} dk.$$

If there exists one k such that the corresponding imaginary part of $\omega(k)$ is strictly positive then the perturbation is growing and the equilibrium is unstable.

Replacing ζ by $\exp(ikx - i\omega t)$ in (5.9) leads to

$$\tau(\omega - k\lambda_1)(\omega - k\lambda_2) + i(\omega - kc) = 0$$

where c, λ_1 and λ_2 are real since the system (5.6) is hyperbolic. For localised perturbation, i.e. small length wave, $|k\lambda_1\tau| \gg 1$ and $\omega(k)$ can be approximated (see [60])

$$\omega \approx \begin{cases} k\lambda_1 - \frac{i}{\tau}\frac{\lambda_1 - c}{\lambda_1 - \lambda_2}, \\ k\lambda_2 - \frac{i}{\tau}\frac{c - \lambda_2}{\lambda_1 - \lambda_2}. \end{cases}$$

To ensure that the imaginary parts of all $\omega(k)$ remain strictly negative, the following conditions must be satisfied:

$$\lambda_1 < c < \lambda_2. \tag{5.10}$$

These conditions are called *sub-characteristic conditions*. If one of these conditions is not fulfilled then the perturbation has locally an exponential growth and the equilibrium is unstable.

In the specific case of the AR model

$$\begin{aligned} \lambda_1 &= v_e - \rho_e p'(\rho_e) \\ \lambda_2 &= v_e \\ c &= v_e - \rho_e V'(\rho_e) \end{aligned}$$

and the conditions (5.10) become

$$-p'(\rho_e) < V'(\rho_e) < 0. \tag{5.11}$$

The second inequality is usually satisfied. For some choices of the function $p(\rho)$, the first inequality can be not satisfied for a range of density values. For such $p(\rho)$ function, equilibria in that range of values are unstable.

There is an easy interpretation of the condition (5.10) as a competition between a “relaxation” and an “anticipation” process:

- The relaxation process is linked to the value of V' and tends to destabilise the system. If the drivers enter a region with an increase of density $\Delta\rho(x) > 0$, their preferential speed $V(\rho)$ decreases. The amplitude of this decrease is linked to the value of $V'\Delta\rho$. Since the drivers are breaking, the density increases in the region which leads to a new decrease of the preferential speed ...

- The anticipation process is linked to the value of $p'(\rho)$ and tends to stabilise the system. The equation (5.2) describing the evolution of the speed using Lagrangian derivative expresses the fact that the drivers anticipate the change of density by adapting their speed in function of $p(\rho)$. Since the drivers break before entering into the perturbation, this tends to spread out this perturbation.

- If the two processes are present, there is a competition and the anticipation process is sufficient to stabilise the system if (5.10) is satisfied.

If the equilibrium is unstable then a small perturbation may grow to a complete traffic jam. For example consider the following $p(\rho)$ function

$$p(\rho) = 35\ln\left(\frac{\rho}{180}\right)$$

together with the preferential speed $V(\rho)$ represented in Fig. 5.2. In that case, equilibria with density in the range 27–65 [veh/km] are unstable (see Fig. 5.3).

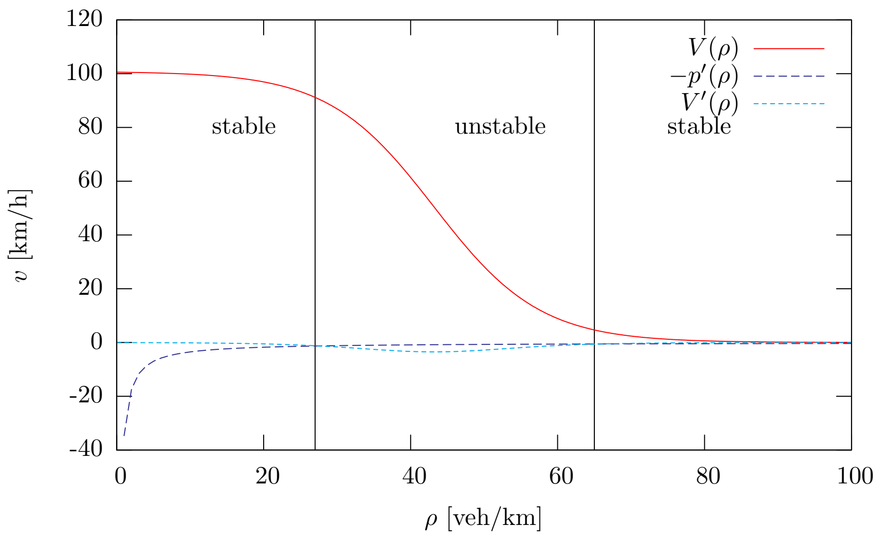


Figure 5.2: For some $p(\rho)$ functions there is a range of density values where the equilibria are unstable.

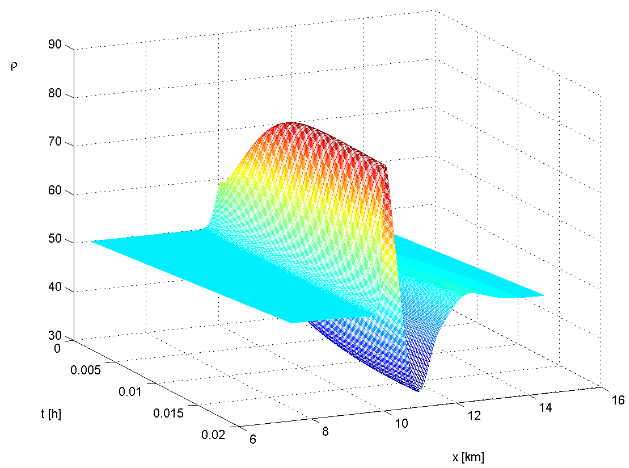


Figure 5.3: The evolution of a small perturbation predicted by the AR model when the equilibrium is unstable.

5.3 Modelling of the junctions

As for the LWR model, a description of the driver's behaviour at the junction must be provided in order to simulate the traffic evolution on a road network. This description is done, like for the LWR model, by expressing the solutions of the Riemann problems at the junctions.

$$(\rho_i(x,0), v_i(x,0)) = (\rho_{i,0}, v_{i,0}) \qquad \forall x,\ \forall i.$$

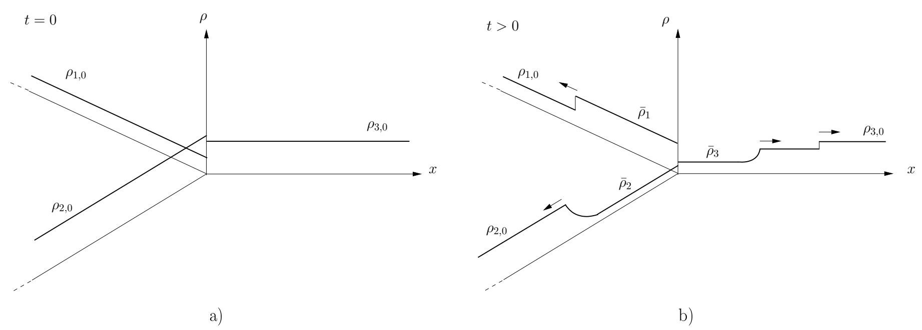


Figure 5.4: A Riemann problem for a junction with two incoming and one outgoing roads: the initial density and the density after some times. The black arrows indicate the position and direction of the different waves.

This section is structured as follow:

1. The admissible regions for the new states near the junction are described.

2. Additional conditions to select a unique condition are introduced.

3. A reformulation of the model using a sending–receiving approach is performed in order to highlight the similarity with [50].

4. The properties of the junction model are analysed in order to validate the introduction of the new assumptions.

5. The stability of equilibria at the junction is analysed allowing to prove, in some cases, the existence of a solution for a network.

6. The existence of a solution in the BV space is discussed.

7. This junction model is compared with other models present in the literature.

5.3.1 The admissible regions for the junction models

As mentioned in the introduction, in order to formulate a junction model the admissible regions for the values of $(\bar{\rho}_i, \bar{v}_i)$ must first be clarified for each road. The shape of this admissible region will be different if an incoming or an outgoing road is considered.

Incoming road

On incoming roads, only the negative waves produced by the Riemann problem

$$(\rho_i(x,0), v_i(x,0)) = \begin{cases} (\bar{\rho}_i, \bar{v}_i) & \text{if } x = 0 \\ (\rho_{i,0}, v_{i,0}) & \text{if } x < 0 \end{cases}$$

are obviously admissible. Since the speed of the second type wave (contact discontinuity with a speed equal to the speed of the drivers) is necessarily positive, the only admissible wave is a wave of the first type (shock or rarefaction wave). Hence $(\bar{\rho}_i, \bar{\rho}_i\bar{v}_i)$ must be on the curve Υ passing through $(\rho_{i,0}, \rho_{i,0}v_{i,0})$ where Υ is defined as

$$\Upsilon = (\rho, \gamma(\rho)) = (\rho, \rho K - \rho p(\rho))$$

with

$$K = v_{i,0} + p(\rho_{i,0}).$$

In this case, the intermediate state (U_c in Figure 5.1) of the Riemann problem $(\rho_{i,0}, \rho_{i,0}v_{i,0})$–$(\bar{\rho}_i, \bar{\rho}_i\bar{v}_i)$ is $(\bar{\rho}_i, \bar{\rho}_i\bar{v}_i)$ and there is no second wave with positive speed. Defining σ as

$$\sigma = \arg\max_{\rho} \gamma(\rho),$$

two cases can be considered:

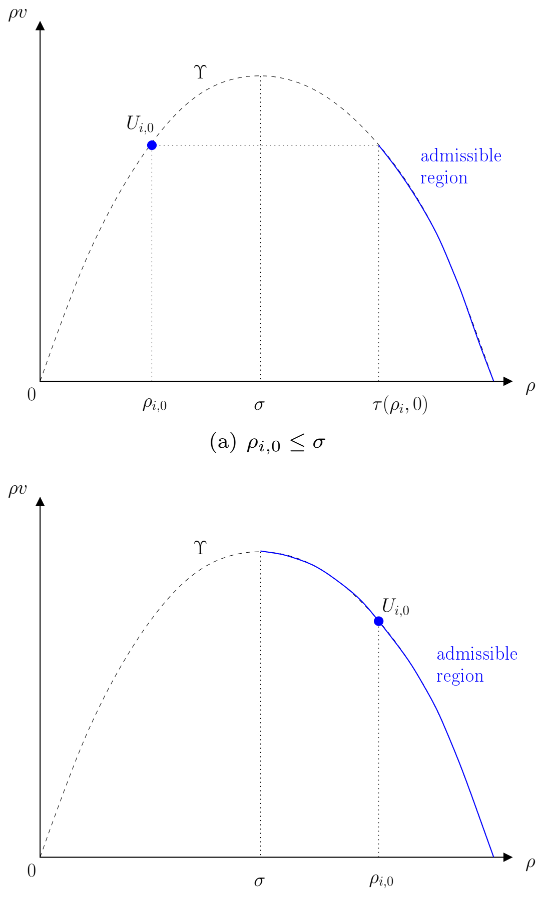
1. If $\rho_{i,0} < \sigma$ then the only possibility to have a wave with negative speed is to have $\bar{\rho}_i > \tau(\rho_{i,0})$ (see Fig. 5.5(a)) where, for each $\rho \neq \sigma$, $\tau(\rho)$ is the unique number $\tau(\rho) \neq \rho$ such that

   $$\gamma(\rho) = \gamma(\tau(\rho)).$$

   In that case, the wave on the incoming road is a shock wave with a negative speed equal to the slope of the curve connecting $U_{i,0}$ to $\bar{U}_i$.

2. If $\rho_{i,0} \geq \sigma$ then two sub-cases must be distinguished:

   - if $\bar{\rho}_i \geq \rho_{i,0}$ the solution consists of a shock wave with a negative speed;
   - if $\bar{\rho}_i \leq \rho_{i,0}$ the solution will be a rarefaction wave. In order that the right limit of this rarefaction wave has a negative speed, it needs that $\bar{\rho}_i \geq \sigma$ (see Fig. 5.5(b)).



(a) $\rho_{i,0} \leq \sigma$

(b) $\rho_{i,0} > \sigma$

Figure 5.5: The admissible regions for an incoming road.

The admissible region for an incoming road is thus composed of the part of the curve Υ represented in Fig. 5.5 and, of course, $(\rho_{i,0}, v_{i,0})$ for which there is no wave.

To be complete, the admissible region when a road is empty ($\rho_{i,0} = 0$) must be specified. In this case, the variable $v_{i,0}$ has no physical meaning but only a mathematical one. Two cases can be considered:

- If $v_{i,0} \neq 0$, one can show that the state $(0, v_{i,0})$ cannot be connected to another state with a non-positive speed. The admissible region is thus $(\rho_{i,0}, v_{i,0})$.

- If $v_{i,0} = 0$, the state $(0, 0)$ can be connected to any state $(\bar{\rho}_i, 0)$. The choice of the density $\bar{\rho}_i$ is arbitrary since all the solutions of the Riemann problem

$$(\rho(x,0), v(x,0)) = \begin{cases} (0,0) & \text{if } x < 0 \\ (\bar{\rho}_i, 0) & \text{if } x = 0 \end{cases}$$

are identical in $L^1[-\infty, 0]$. Even if all values are possible for $\bar{\rho}_i$, the most meaningful is zero like the density on the incoming road.

It can be noticed a great similarity with the LWR first order models for which a value called the "sending capacity" or the "traffic demand" (3.8) was introduced. This value represents the greatest possible outflow of a road segment and it was shown equal to

$$\text{sending capacity} = \begin{cases} f(\rho) & \text{if } \rho \leq \arg\max_\rho f(\rho) \\ \max f(\rho) & \text{if } \rho > \arg\max_\rho f(\rho) \end{cases}$$

where $f(\rho)$ represents the flow associated to the density ρ. In this second order model, the greatest possible outflow of a road segment is the maximal flow for a point on the admissible region and is equal to

$$\text{sending capacity} = \begin{cases} \gamma(\rho) & \text{if } \rho \leq \arg\max_\rho \gamma(\rho) \\ \max \gamma(\rho) & \text{if } \rho > \arg\max_\rho \gamma(\rho). \end{cases} \tag{5.12}$$

The similarity between the two models is obvious with the replacement of $f(\rho)$ by $\gamma(\rho)$. In the AR model, the sending capacity is function of the density but also of the speed (via the value of $v_{i,0} + p(\rho_{i,0})$ which is used in the definition of γ).

The admissible region satisfies some intuitive ideas:

- if there is nobody on the incoming road ($\rho_{i,0} = 0$), the maximal flow allowed to leave the road is zero;
- if there are few vehicles on the incoming road ($\rho_{i,0} \ll \sigma$), the flow allowed to leave the road is low (less than $\rho_{i,0}v_{i,0}$);
- if there is a lot of vehicles on the incoming road ($\rho_{i,0} \gg \sigma$), the flow allowed to leave the road may be high (up to $\gamma(\sigma)$).

Outgoing road

On outgoing roads, only the positive waves produced by the Riemann problem

$$(\rho_i(x,0), v_i(x,0)) = \begin{cases} (\bar{\rho}_i, \bar{v}_i) & \text{if } x = 0 \\ (\rho_{i,0}, v_{i,0}) & \text{if } x > 0 \end{cases}$$

are obviously admissible. This case is more complex than the previous one since the two waves may be simultaneous present. The second wave having always a positive speed, any intermediate state U_c on the curve $(\rho, \rho v_{i,0})$ can be connected to $U_{i,0} = (\rho_{i,0}, \rho_{i,0}v_{i,0})$. The admissible region is thus the $(\rho, \rho v_{i,0})$ curve and the region of the $(\rho, \rho v)$ plane which can be connected to the curve $(\rho, \rho v_{i,0})$ with any rarefaction or shock wave whose speed is positive.

In order to characterise the admissible region, all the possible curves Υ are considered. Υ has been defined as a curve passing by the left state of the Riemann problem ($(\bar{\rho}_i, \bar{\rho}_i\bar{v}_i)$ in this case). Since Υ only depends on $K = \bar{v}_i + p(\bar{\rho}_i)$, considering all curves Υ is the same as considering all the possible values of $K = \bar{v}_i + p(\bar{\rho}_i) \in]0, \infty[$. To emphasise the fact that the curve Υ depends on the value K, it will be denoted Υ_K from now on.

Three cases must be considered (see Fig. 5.6):

a) If $0 \leq K \leq v_{i,0}$ then the curve Υ_K is necessarily located under the curve $(\rho, \rho v_{i,0})$. The left state $(\bar{\rho}_i, \bar{v}_i)$ is connected to the intermediate state U_c, which is the vacuum $(\rho = 0)$, by a rarefaction wave. In order that the left limit of this rarefaction wave has a positive speed, it is needed that

$$\left.\frac{d(\rho K - \rho p(\rho))}{d\rho}\right|_{\rho=\bar{\rho}_i} \geq 0.$$

It implies that $\bar{\rho}_i$ must be lesser than σ_K (see Fig. 5.6(a)).

b) If $v_{i,0} < K$ and $\gamma_K(\sigma_K) \leq \sigma_K v_{i,0}$ then the curve Υ_K is greater than the line $(\rho, \rho v_{i,0})$ for small densities but crosses this line before its maximum. If $(\bar{\rho}_i, \bar{\rho}_i\bar{v}_i)$ is on the part of Υ_K which is above the curve $(\rho, \rho v_{i,0})$, it is connected to the intermediate state U_c by a shock wave with positive speed. If $(\bar{\rho}_i, \bar{\rho}_i\bar{v}_i)$ is on the part of Υ_K which is under the curve $(\rho, \rho v_{i,0})$, it is connected to U_c by a rarefaction wave. In order that the left limit of this rarefaction wave has a positive speed, $\bar{\rho}_i$ must be lesser than σ_K (see Fig. 5.6(b)).

c) If $v_{i,0} < K$ and $\gamma_K(\sigma_K) > \sigma_K v_{i,0}$ then the maximum of Υ_K is above the line $(\rho, \rho v_{i,0})$. If $(\bar{\rho}_i, \bar{\rho}_i\bar{v}_i)$ is on the part of Υ_K which is under the curve $(\rho, \rho v_{i,0})$, it should be connected to U_c by a rarefaction wave with a negative speed, which is impossible.

If $(\bar{\rho}_i, \bar{\rho}_i\bar{v}_i)$ is above the straight line, it is connected to U_c by a shock wave whose speed is the slope of the curve connecting $(\bar{\rho}_i, \bar{\rho}_i\bar{v}_i)$ to U_c. In order to have a positive slope, it needs that

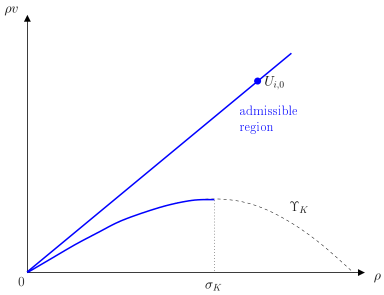
$$\bar{\rho}_i \leq \tau_K(\rho_c)$$

where ρ_c is the density of the intermediate state U_c (see Fig. 5.6(c)).

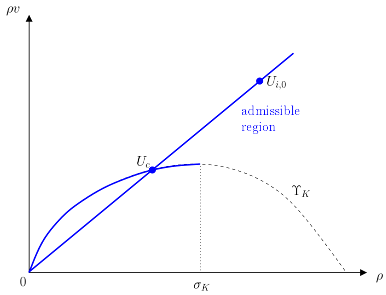
If the admissible regions for all the values of K are combined, the global admissible region for an outgoing road can be obtained (see Fig 5.7).

In fact, the admissible region only depends on the speed of the vehicles on the road. If $v_1 \leq v_2$, the admissible region associated with v_1 will be included in the admissible region associated with v_2.

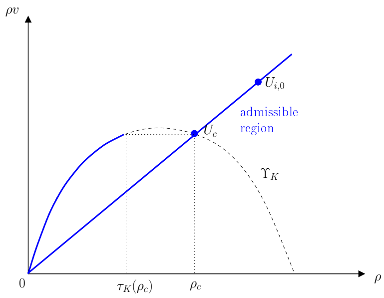
Two limit cases can also be considered:



(a) $0 \le K \le v_{i,0}$



(b) $v_{i,0} < K$ & $\gamma_K(\sigma_K) \le \sigma_K v_{i,0}$



(c) $v_{i,0} < K$ & $\gamma_K(\sigma_K) > \sigma_K v_{i,0}$

Figure 5.6: Some parts of the admissible region for an outgoing road.

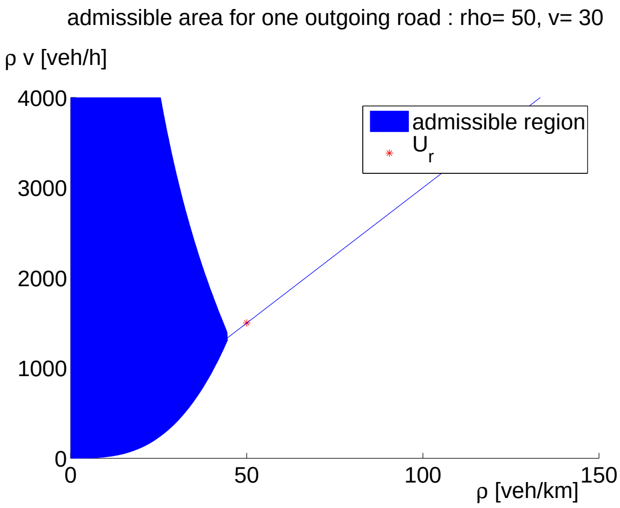


Figure 5.7: The admissible region for an outgoing road.

- If $v_{i,0} = 0$, then the admissible region is reduced to the states $(\bar{\rho}_i, 0)$ for any $\bar{\rho}_i$. The choice of the density $\bar{\rho}_i$ is arbitrary since all the solutions of the Riemann problem on the outgoing road are identical in $L^1[0, \infty]$. Even if all values are possible for $\bar{\rho}_i$, the most meaningful is $\rho_{i,0}$ i.e. the density on the incoming road.

- If $\rho_{i,0} = 0$, then the admissible region is simply function of the speed $v_{i,0}$ as described above. Of course, the variable $v_{i,0}$ has no more any physical meaning. To remove the inconsistence of drivers acting in function of a non physical variable, it may be useful to add a relaxation term $\frac{V(\rho)-v}{\tau}$ at the right of (5.2). This relaxation term tends to remove such situations where $\rho_{i,0}$ and $v_{i,0}$ are both equal to zero.

This admissible region satisfies some intuitive ideas:

- if there is almost nobody on the outgoing road ($\rho_{i,0} \approx 0$, $v_{i,0} \gg 0$), the admissible region is quite large;

- if there are many vehicles on the outgoing road ($\rho_{i,0} \gg 0$, $v_{i,0} \approx 0$), the admissible region is smaller.

5.3.2 Additional conditions

The description of the admissible regions has shown that many states are admissible. It is clear that additional conditions are needed in order to have a unique solution to the Riemann problem. The choice of these conditions is the main part of the establishment of a new road junction model.

1. A first condition is indisputable: the conservation of flow. The sum of the entering flows must be equal to the sum of the leaving flows at the junction.

2. The second equation of the AR model (5.2) describes the behaviour of the drivers. It says that the Lagrangian derivative of the speed is equal to the Lagrangian derivative of $-p(\rho)$. It means that a driver will adapt his speed when the quantity $p(\rho)$ is modified.

   The most natural extension of this behaviour at the junction is

   $$v_a - v_b = p(\rho_a) - p(\rho_b)$$

   where the subscripts a and b means "after" and "before" the junction. In other words, the quantity $v + p(\rho)$, which describes the behaviour of the drivers, is "conserved" through the junction. Here, the meaning of conservation is not the same as in "conservation of the flow": the total flow of the quantity $v + p(\rho)$ is not necessarily the same before and after the junction but each driver tends to conserve his quantity $v + p(\rho)$ which describes his behaviour. If there is only one incoming road, all the drivers have the same behaviour, and thus:

$$v_a + p(\rho_a) = v_b + p(\rho_b).$$

In the case where there are several incoming roads, it is natural to consider that the behaviour of the drivers after the junction is an average of the driver behaviours from the incoming roads. The additional condition used in our model to represent this fact is:

$$v_a + p(\rho_a) = \sum_i \alpha_i (v_i + p(\rho_i)) \tag{5.13}$$

where α_i is the proportion of the drivers coming from the incoming road i.

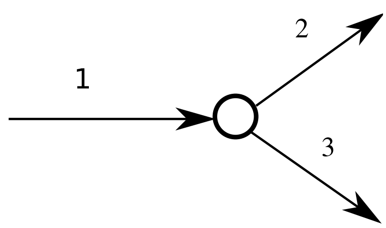
It must be noticed that (5.13) is an additional *assumption* which is not motivated by microscopic considerations. The plausibility of this assumption is confirmed by the properties of the resulting model (see Section 5.3.4).

Another approach could be to return to the microscopic model underlying the Aw & Rascle model (see [1]) and analyse what happens at the junction. This approach is taken in [32, 33] and leads to a homogenisation problem which needs to introduce a new function $p(\rho)$ after the junction.

3. The two previous additional conditions are not sufficient to have a unique solution to the Riemann problem at the junction. In order to get a unique solution, like for the LWR model (see [35], [8], [43] and [38]), it may be assumed that the drivers behave in such a way that the flow entering the outgoing roads is maximised with respect to the previous restrictions.

4. In some cases, there still exist two solutions and the one with maximal speed may be privileged. This is reasonable since the drivers tend to act in order to maximise their speed. In fact, it does not matter as the solution will be the same in the L^2 norm whatsoever the choice.

The diverging junction



It is reasonable to assume that, the drivers having fixed destination aims, the proportions of the total flow entering into road 2 and 3 are fixed (α_2 and α_3).

With the additional conditions presented in section 5.3.2, the optimisation problem at the junction can be expressed as

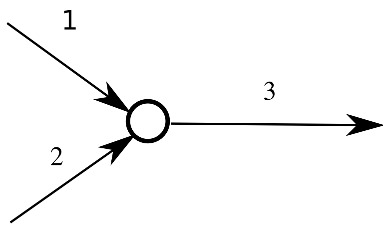
$$\max_{\bar{\rho}_i,\bar{v}_i} \bar{\rho}_1\bar{v}_1$$

subject to

$$\left\{\begin{array}{rl} \bar{\rho}_1\bar{v}_1 = & \frac{\bar{\rho}_2\bar{v}_2}{\alpha_2} = \frac{\bar{\rho}_3\bar{v}_3}{\alpha_3} \\ \bar{v}_1 + p(\bar{\rho}_1) = & \bar{v}_2 + p(\bar{\rho}_2) = \bar{v}_3 + p(\bar{\rho}_3) \\ (\bar{\rho}_i,\bar{v}_i) \in & \text{admissible region of road } i. \end{array}\right. \tag{5.14}$$

The solution of this optimisation problem produces the new values of (ρ_i, v_i) at the junctions which allow to solve the Riemann problems on each road.

The merge junction



For the merge junction, one need to introduce a coefficient describing how the available space on the outgoing road is shared by the incoming roads in case of congestion. A simple way to introduce this coefficient is to make it depend on the flows wishing to enter the outgoing road (see [38]):

$$\begin{aligned} \alpha_1 &= \frac{f_1^*}{f_1^* + f_2^*} \\ \alpha_2 &= 1 - \alpha_1 \end{aligned} \tag{5.15}$$

where f_i^* is the maximal flow able to leave the road i (the "sending capacity"). Other coefficients may also be considered. With the additional conditions introduced above, the optimisation problem at the junction can be expressed as

$$\max_{\bar{\rho}_i,\bar{v}_i} \bar{\rho}_1\bar{v}_1 + \bar{\rho}_2\bar{v}_2$$

subject to

$$\left\{\begin{array}{rl} \frac{\bar{\rho}_1\bar{v}_1}{\alpha_1} = & \frac{\bar{\rho}_2\bar{v}_2}{\alpha_2} = \bar{\rho}_3\bar{v}_3 \\ \bar{v}_3 + p(\bar{\rho}_3) = & \alpha_1(\bar{v}_1 + p(\bar{\rho}_1)) + \alpha_2(\bar{v}_2 + p(\bar{\rho}_2)) \\ (\bar{\rho}_i,\bar{v}_i) \in & \text{admissible region of road } i. \end{array}\right.$$

Solving this optimisation problem produces the new values of (ρ_i, v_i) at the junctions. These new values are used as initial conditions for the Riemann problems on each road.

5.3.3 Reformulation using a sending–receiving approach

As explained in the introduction this junction model turns out to be very similar to the model [50] even if the formulations appear different. To emphasise this similarity our model can be reformulated using a sending–receiving terminology.

For simplicity consider an 1-incoming-1-outgoing road junction with initial condition such that

$$v_{1,0} + p(\rho_{1,0}) = K_1.$$

Our model leads to

$$\begin{aligned} \max \quad & \bar\rho_1 \bar v_1 && (5.16)\\ (\bar\rho_1, \bar v_1) \quad & \in \quad AR_1 && (5.17)\\ (\bar\rho_2, \bar v_2) \quad & \in \quad AR_2 && (5.18)\\ \bar\rho_1 \bar v_1 \quad & = \quad \bar\rho_2 \bar v_2 && (5.19)\\ \bar v_1 + p(\bar\rho_1) \quad & = \quad \bar v_2 + p(\bar\rho_2). && (5.20) \end{aligned}$$

The first condition $(\bar\rho_1, \bar v_1) \in AR_1$ implies (see Fig. 5.5) that

$$\bar q_1 = \bar\rho_1 \bar v_1 \le S(\rho_{1,0}, K_1) \tag{5.21}$$

where $S(\rho, K)$ is defined as (5.12). The similarity with the LWR model is obvious except that the sending capacity is now also function of K_1. The function f used in the LWR model to define the "sending capacity" (see (3.8) at page 31) is replaced by

$$\gamma(\rho) = -\rho p(\rho) + K\rho.$$

If the particular function $p(\rho) = -V(\rho)$ is used then the similarity is more obvious since

$$\gamma(\rho) = \rho V(\rho) + K\rho.$$

It means that the function used is the classical flow function shifted proportionally to $K = v - V(\rho)$, the difference between the drivers' speed and the preferential speed.

The first condition (5.17) also implies that $\bar v_1 + p(\bar\rho_1) = K_1$. Combined with (5.20) it leads to $\bar v_2 + p(\bar\rho_2) = K_1$

The second condition $(\bar\rho_2, \bar v_2) \in AR_2$ combined with $\bar v_2 + p(\bar\rho_2) = K_1$ implies (see Fig. 5.6) that

$$\bar q_2 = \bar\rho_2 \bar v_2 \le R(\rho_c, K_1) \tag{5.22}$$

where the receiving capacity R is defined as

$$R(\rho, K) = \begin{cases} \max \gamma(\rho) & \text{if } \rho \le \arg\max_\rho \gamma(\rho) \\ \gamma(\rho) & \text{if } \rho > \max_\rho \gamma(\rho). \end{cases}$$

and

$$\begin{aligned} \rho_c \quad & = \quad p^{-1}(K1 - v_2) \\ \gamma(\rho) \quad & = \quad -\rho p(\rho) + K\rho. \end{aligned}$$

The similarity with the receiving capacity defined for the LWR model (see (3.9) page 32) is also obvious.

Combining (5.16), (5.21) and (5.22), one has

$$\bar{q}_1 = \min\big(S(\rho_{1,0}, K_1), R(\rho_c, K_1)\big).$$

This equation expresses that the total passing flow is the minimum of the sending capacity of the incoming road and the receiving capacity of the outgoing road. This equation appears very similar to (3.13)

With this formulation it is easy to see that our model and [50] are very similar. For example the problem of a diverging junction is formulated in [50] as

$$\begin{aligned} \bar{q}_1 &= \min\big(\alpha_1 S(\rho_{1,0}, K_1), R(\rho_{c2}, K_1)\big) \\ \bar{q}_2 &= \min\big(\alpha_2 S(\rho_{1,0}, K_1), R(\rho_{c3}, K_1)\big) \end{aligned}$$

which is exactly the reformulation of (5.14) in terms of receiving and sending capacities.

The only slight differences between our model and [50] are that the sharing coefficients for a merge junction are fixed in [50] and that the particular function $p(\rho) = -V(\rho)$ is used. A more general function is however considered in [41].

5.3.4 Properties of the junction model

The solutions of the Riemann problem at the junctions have some interesting properties that are described in this section. It is successively shown that the solution of the optimisation exists and is unique (taking into account Condition 4 of Section 5.3.2), that the solutions are physically acceptable and finally that they are able to represent some interesting phenomena such as the *capacity drop phenomenon*.

Existence and uniqueness

The existence and uniqueness is proved by the following reasoning:

1. Given the initial states on the incoming roads, it is possible to compute the values $(v_{i,0} + p(\rho_{i,0}))$, the curves Υ_{K_i} and thus the flow which can leave the incoming roads $[0, f_i^*]$.
2. Given the values $(v_{i,0} + p(\rho_{i,0}))$ on the incoming roads, the curves Υ_{K_j} to which the new states of the outgoing roads belong can be computed.
3. On an outgoing road j, there always exists an intersection between the curve Υ_{K_j} and the admissible region (at least the state $(0,0)$). Once the left part of the curve Υ_{K_j} enters the admissible region of road j, it stays inside the admissible region for all lower density/flow on Υ_{K_j}. So there exists a maximal flow f_j^* such that all flows in $[0, f_j^*]$ are feasible with a state in the admissible region and on Υ_{K_j}.

4. The problem of maximising the passing flow on a merge junction in terms of flows

$$\max \bar{f}_3$$

subject to

$$\begin{aligned} \bar{f}_1 + \bar{f}_2 &= \bar{f}_3 \\ \bar{f}_1 &= \alpha_1 \bar{f}_3 \\ \bar{f}_2 &= \alpha_2 \bar{f}_3 \\ \bar{f}_1 &\in [0, f_1^*] \\ \bar{f}_2 &\in [0, f_2^*] \\ \bar{f}_3 &\in [0, f_3^*] \end{aligned}$$

has a unique solution.

5. The problem of maximising the passing flow on an diverging junction in terms of flows

$$\max \bar{f}_1$$

subject to

$$\begin{aligned} \bar{f}_1 &= \bar{f}_2 + \bar{f}_3 \\ \alpha_2 \bar{f}_1 &= \bar{f}_2 \\ \alpha_3 \bar{f}_1 &= \bar{f}_3 \\ \bar{f}_1 &\in [0, f_1^*] \\ \bar{f}_2 &\in [0, f_2^*] \\ \bar{f}_3 &\in [0, f_3^*] \end{aligned}$$

has a unique solution.

6. There is a unique correspondence between a solution in terms of flows and a solution in terms of density/speed. In general, for an incoming road there is only one point of the admissible region which corresponds to a specified flow and, for an outgoing road, only one point of Υ_{K_j} inside the admissible region and corresponding to a specified flow.

   In some particular cases, there may be two possible points in the admissible region. This may happen on an incoming road if $\bar{f}_1 = f_1^*$ and $\rho_{1,0} < \sigma_K$ and on an outgoing road if the point at the intersection between Υ_{K_j} and the line $\rho v = \rho v_{j,0}$ correspond to the specified flow. For simplicity one can take as final solution the states corresponding to the maximal speed. This choice between the two possible points has no real consequence since the two lead to the same solution in the L^2 space.

Physical acceptability

The solutions of this junction model are intuitively acceptable.

A first important property of this junction model is his coherence with the AR single road model. For the diverging junction if all the incoming drivers take the same outgoing road ($\alpha_2 = 0, \alpha_3 = 1$) or for the merge junction if one of the incoming roads is empty ($\rho = 0$), the solutions are the same as the classical Aw & Rascle model for an infinite single road.

A second important point is that numerical simulations produce plausible results. Consider the merge junction case. If there is not too much traffic on the roads, all the flow on the incoming roads may pass (see Table 5.1). In this situation the total outflow in the final situation (2880 veh/h) is equal to the sum of the initial inflows (2×1440 veh/h). In the second simulation (see Table 5.2) the traffic on the incoming roads is more important. In this case only a part (2165 veh/h) of the total inflow flow (3300 veh/h) may pass.

The numerical results presented in the tables were performed with the $p(\rho)$ function defined as

$$p(\rho) = \frac{v_{ref}}{\gamma}\left(\frac{\rho}{\rho_{max}}\right)^{\gamma}$$

with $v_{ref} = 120$ [km/h], $\rho_{max} = 90$ [veh/km] and $\gamma = 2$. This is a plausible $p(\rho)$ function derived from microscopic consideration (see [1]).

	Incoming roads			Outgoing road		
	ρ	v	f	ρ	v	f
Initial states	20	72	1440	51.4	58.36	3000
	20	72	1440			
Final states	20	72	1440	53.9	53.43	2880
	20	72	1440			

Table 5.1: There is not too much traffic on the two incoming roads for the capacity of the outgoing road. The whole flow may pass.

	Incoming roads			Outgoing road		
	ρ	v	f	ρ	v	f
Initial states	30	55	1650	51.4	58.36	3000
	30	55	1650			
Final states	80.7	13.42	1082	52	41	2165
	80.7	13.42	1082			

Table 5.2: There is too much traffic on the two incoming roads for the capacity of the outgoing road. Only a fraction of the flow may pass.

A last point is the analysis of the situations where no flow can pass through the junction. For a merge junction, this situation only happens when there is no traffic on the incoming road or when the speed on the outgoing road is equal to zero (the admissible region is reduced to curve $(\rho, 0)$ which is associated with a zero flow). For the diverging junction, the only situation where no flow can leave the incoming road is when the speed on one of the outgoing roads is equal to zero (the admissible region is reduced to curve $(\rho, 0)$ which is associated with a zero flow).

This implies that, if only one of the outgoing roads is jammed, the outflow of the other road is equal to zero. This is a rather hard constraint on the outflows. The reason is that the AR model, used to describe the evolution on one road, does not make any distinction between the different lanes. In a single lane road, if a driver stops because he is unable to turn left, he also blocks all the drivers wanting to turn right. To remove this hard constraint, it would be needed to adapt the AR model to take the different lanes into account.

Invariant region

The single road model (5.1)–(5.2) admits the invariant region (5.4)

$$\mathcal{R} := \{(\rho, v) \mid 0 \leq v + p(\rho) \leq v_m, \quad \rho \geq 0,\ 0 \leq v \leq v_m\}.$$

It is easy to see that the new proposed model for the junction keeps this invariant region. The first inequalities are satisfied since a new value of $v + p(\rho)$ may only appear on an outgoing road and this new value is either equal to the value on an incoming road or a convex combination of the values on two incoming roads.

The solution of the maximisation problem always exists with $\rho \geq 0$ and $v \geq 0$ and, since $p(\rho) \geq 0$, the last inequalities are trivial.

Capacity drop phenomenon

The proposed junction model might be able to naturally represent the *capacity drop phenomenon* which has been introduced in Section 3.2.3. The reason why this new junction model is able to represent this phenomenon is illustrated in Figure 5.8. In this figure are represented the graphs relative to one of the two incoming roads and to the outgoing one. An initial equilibrium (states 1 of Fig. 5.8) where the state of the outgoing road is on the border of the admissible region is considered. Starting from the initial state 1, one considers on the incoming road an increase of flow ($\rho v \nearrow$) with an increase of density and a decrease of speed ($\rho \nearrow, v \searrow$) such that the quantity $v + p(\rho)$ decreases (state 2). A new Riemann problem at the junction must now be solved. Since the quantity $v + p(\rho)$ has decreased on one of the incoming road, the new state on the outgoing road must be on a curve $v + p(\rho) = C_2^{out}$ lower than the curve $v + p(\rho) = C_1^{out}$. Since the total flow able to leave the incoming road is greater than the initial flow (f_1^{in}), the new state on the outgoing road will be state 3 (the state with the maximum flow on the curve $\Upsilon_{C_2^{out}}$). Assuming that nothing has changed on the second incoming road, a decrease of the flow on the outgoing road implies a decrease of the flow on the incoming roads. The new flow of the incoming road must be less than f_1^{in} (for example state 3).

To illustrate this phenomenon, a series of simulations has been carried out. The corresponding solutions are presented in Table 5.3. To simplify, a case

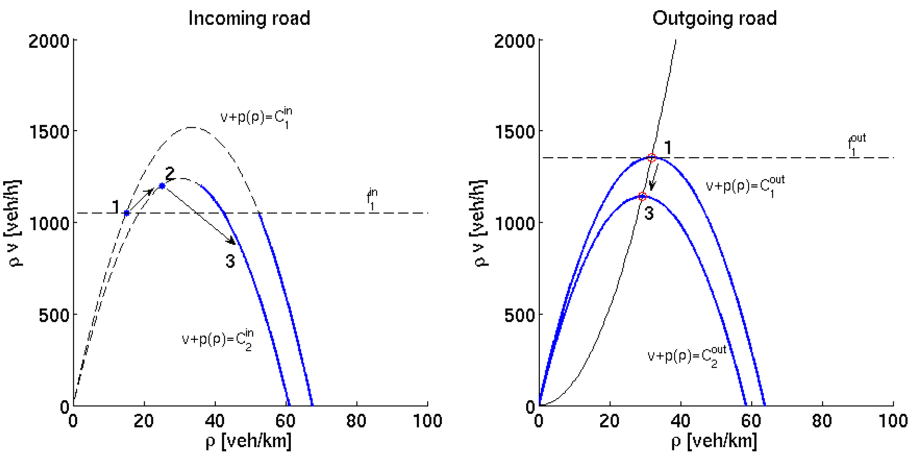


Figure 5.8: Illustration of the capacity drop phenomenon with an increase of flow at one incoming road.

where the states on road one and road two are the same is considered. Starting from an equilibrium (row 1) with a total passing flow of 3000 veh/h, a perturbation (row 2) is applied. This perturbation consists of an increase of the incoming flow (3300 veh/h) and lead to the new equilibrium represented in row 3. To this new equilibrium corresponds a total passing flow of 2165 veh/h. This illustrates the capacity drop phenomenon: the incoming flows are higher but the outgoing flow drops down. Moreover, if one tries to return to the initial state by an opposite perturbation on the incoming roads (row 4), the initial total passing flow is not recovered, but the resulting flow is a lower one.

	state	ρ_1	v_1	f_1	ρ_3	v_3	f_3
1	equilibrium	20	75	1500	51.4	58.36	3000
2	perturbation	30	55	1650	51.4	58.36	3000
3	equilibrium	80.7	13.42	1082.5	52	41.6	2165
4	perturbation	20	75	1500	52	41.6	2165
5	equilibrium	91.5	15.9	1458.5	47.8	61	2917

Table 5.3: The capacity drop phenomenon illustrated in a particular case.

To summarise, this increase of inflows leading to a decrease of outflow can only appear if:

- the state on the outgoing road remains on the boundary of the admissible region;
- the inflows are such that the mean "behaviour" of the drivers ($\sum_i \beta_i(v_i + p(\rho_i))$) decreases.

The total decrease of the outflow is linked to the value of the decrease of $v+p(\rho)$ for a fixed increase of the flow ($\rho v \nearrow, \rho \nearrow, v \searrow$). So it can be claimed that the importance of the capacity drop phenomenon is linked to the value of $p'(\rho)$. If the derivative is small then the phenomenon may be important. To the author's knowledge, this possibility of increasing inflows leading to a decreasing outflow is not present in the junction models developed so far for the LWR model. These models are thus unable to represent the capacity drop phenomenon.

5.3.5 Stability of equilibria

In this section, the stability of the equilibria is investigated. The notion of stability used here comes from [20]. An equilibrium at a junction is defined as a solution of the Riemann problem (i.e. the (ρ_i, v_i) values at the junction satisfying the conditions of Section 5.3.2). If a wave arrives at the junction, it will perturb the initial state and a new equilibrium will appear. Stability means that small perturbations produce small variations of the equilibrium. More precisely, an equilibrium (ρ_i^*, v_i^*) is *stable* if there exists two constants $C > 0$ and $\delta > 0$ such that:

$$\forall (d\rho, dv) \text{ s.a. } |d\rho| + |dv| < \delta \Rightarrow \sum_i |\rho_i^* - \bar{\rho}_i| + |v_i^* - \bar{v}_i| \leq C(|d\rho| + |dv|) \quad (5.23)$$

where $(d\rho, dv)$ is the perturbation on one of the roads and $(\bar{\rho}_i, \bar{v}_i)$ is the new equilibrium on road i. The equilibrium will be *unstable* if it is impossible to find $C > 0$ and $\delta > 0$ such that (5.23) is true.

To analyse the different possibilities, the curve Γ corresponding to the maxima of Υ_K is introduced. This curve separates the $(\rho, \rho v)$ plane in two regions D_1 and D_2. In Figure 5.9, are represented the different possibilities for the position of the equilibrium states. For an incoming road, the possible region for the equilibria can be decomposed as $\overset{\circ}{D}_1 \bigcup \Gamma \bigcup \overset{\circ}{D}_2$ and for an outgoing road as $\overset{\circ}{D}_1 \bigcup \Gamma$ where the superscript $^\circ$ refers to the interior. In this study of the stability, the particular cases where ρ or v equal zero is excluded.

The three following claims can be proven:

1. If one of the states on the incoming roads belongs to $\overset{\circ}{D}_1$ and one of the states on the outgoing roads belongs to Γ then the equilibrium is unstable.

2. If all the states on the outgoing roads belong to $\overset{\circ}{D}_1$ then the equilibrium is stable.

3. If all the states on the incoming roads belong to $\overset{\circ}{D}_2$ and at most one state on the outgoing roads belongs to Γ then the equilibrium is stable.

During the proofs, the following conventions are taken:

- the superscripts * and $^-$ refer to the initial states before and the final states after the perturbation;

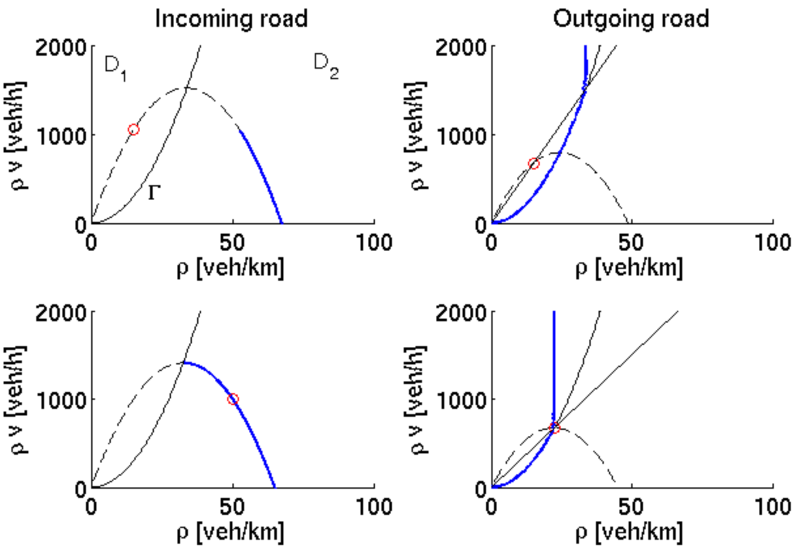


Figure 5.9: The red circles indicate the possible initial equilibrium states and the blue lines are the boundaries of the corresponding admissible regions.

- the subscript i is dropped, ρ^* and $\bar{\rho}$ always refer to the density on a specific road depending on the context;
- $d\rho, df$ and dv refer to the variations of density, flow and speed on the road associated to the perturbation;
- a quantity is said to be bounded when it is always less than $|d\rho| + |dv|$ multiplied by a constant.

The proofs are carried out by showing that, on each road, there exists a constant $C > 0$ (which may be different for each road and function of the equilibrium under consideration) such that $|\rho_i^* - \bar{\rho}_i| + |v_i^* - \bar{v}_i| \leq C(|d\rho| + |dv|)$.

First claim

> If one of the states on the incoming roads belongs to $\overset{\circ}{D}_1$ and one of the states on the outgoing roads belongs to Γ then the equilibrium is unstable.

An example of instability has already been presented in Figure 5.8 for the illustration of the capacity drop phenomenon. In this situation, it is possible, with an arbitrary small change in ρ and ρv on the incoming road, to make the state on this road jump to the region D_2. The variation of the density in the final solution will thus be of the order of $\tau_K(\rho_{i,0}) - \rho_{i,0}$ which is not linked to the amplitude of the perturbation. This equilibrium is unstable.

Second claim

> If all the states on the outgoing roads belong to $\mathring{D}_1$ then the equilibrium is stable.

The proof is divided in two parts. First the influence of a perturbation occurring on an outgoing road is analysed and then the one of a perturbation on an incoming road.

Perturbation on an outgoing road
If the perturbation occurs on an outgoing road, there is a lower bound on the amplitude of this perturbations. This bound comes from the fact that only shock waves can reach the junction from an outgoing road and these shock waves should have a large amplitude in order to make the state on the outgoing road jump from $\mathring{D}_1$ to $\mathring{D}_2$. If it is imposed that δ in (5.23) is smaller than this bound then it is impossible to have a perturbation coming from an outgoing road.

Perturbation on an incoming road
If the perturbation occurs on an incoming road and is small enough, the flow associated with this new state may cross the junction. This state becomes the new equilibrium state on the incoming road in the final solution. The states on the outgoing roads may change but the variation can be bounded by the following four steps:

1. The amplitude of the flow variation can be bounded in function of the density–speed perturbation. Indeed on the incoming road

$$\begin{aligned} f &= \rho v \\ f + df &= (\rho + d\rho)(v + dv) \\ df &= v d\rho + \rho dv + d\rho dv. \end{aligned}$$

   Based on the invariant region $\mathcal{R}$ defined in (5.4), introduce $V = \max(v_m, \rho_m)$ where $\rho_m = p^{-1}(v_m)$. Denoting A the amplitude of the initial perturbation

$$A = |d\rho| + |dv| \,,$$

   if $A \leq 1$ the amplitude of the flow perturbation may be bounded

$$|df| \leq (2V + 1)A = C^1 A.$$

   Since the variation of the flow on the outgoing road is lesser than or equal to $|df|$, the flow variation on the outgoing road is bounded. The flow variation is lesser if this is a diverging junction. In this case, the variation of flow on the incoming road has been shared between the two outgoing roads.

2. The variation of the curve $\Upsilon_{K_{out}}$ to which belongs the state on the outgoing road can be bounded. Indeed, on the incoming road one had previously
$$v + p(\rho) = K_{in},$$
and, after the change, one has
$$v + dv + p(\rho + d\rho) = K'_{in}$$
with
$$\begin{aligned} |K_{in} - K'_{in}| &\leq |dv| + \max(p'(\rho))\,|d\rho| \\ &\leq (1 + \max(p'(\rho)))A = C^2 A. \end{aligned}$$
On the outgoing road the state which was laying previously on the curve $\Upsilon_{K_{out}}$ will now belongs to the curve $\Upsilon_{K'_{out}}$ with $K'_{out} = K_{out} + dK_{out}$. Since K'_{out} is a convex combination of the values of K'_{in} on the incoming roads, one has
$$|dK_{out}| \leq |K_{in} - K'_{in}| \leq C^2 A.$$
The new state on the outgoing road belongs to the curve
$$\Upsilon_{K'_{out}} = \{(\rho, \rho v) \mid \rho v = K_{out}\rho - \rho p(\rho) + dK_{out}\rho\}.$$
On Figure 5.10, the new state will be between the two lines ($\rho v = f^* + |df|$ and $\rho v = f^* - |df|$) and on the curve $\Upsilon_{K_{out}}$ shifted by $dK_{out}\rho$.

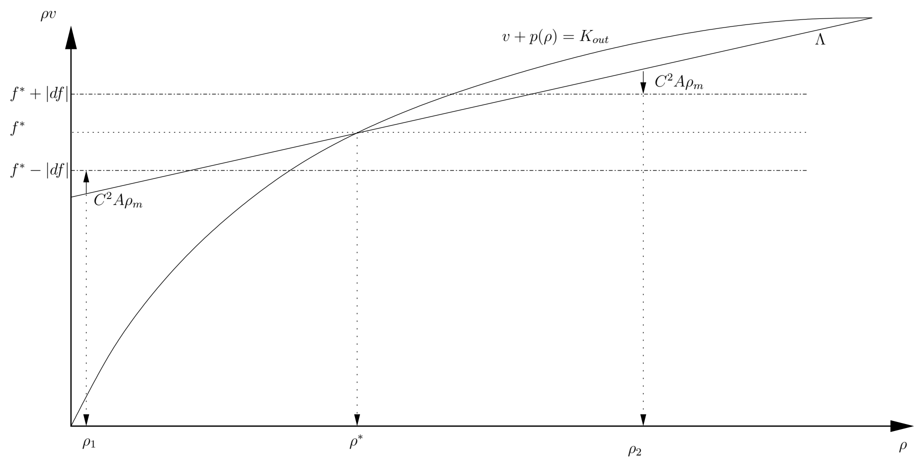


Figure 5.10: Bounds on the variation of ρ and f on the outgoing road. The state (ρ^*, f^*) is the original state before the perturbation.

3. It is possible to find a bound on the new value $\bar{\rho}$ for the outgoing road. Consider the line Λ passing through the original state (ρ^*, f^*) and the maximum of the curve $\Upsilon_{K_{out}}$. Since $(\rho^*, f^*) \in \overset{\circ}{D}_1$ and $(\rho p(\rho))'' > 0$, this

line has a derivative $a > 0$. Assuming that the new curve $\Upsilon_{K'_{out}}$ will be $\Upsilon_{K_{out}}$ shifted by at most $C^2 A\rho_m$ and that A is small enough such that the cross of the lines occur as represented in Figure 5.10 then:

- on the right of ρ^*, the curve $\Upsilon_{K_{out}}$ is above Λ and is shifted less than $C^2\rho_m$ below. The possible intersection between $\Upsilon_{K'_{out}}$ and the line $f^* + |df|$ occurs on the left of ρ_2 with
$$|\rho_2 - \rho^*| = \frac{df + C^2 A\rho_m}{a} \leq \frac{C^1 + C^2\rho_m}{a} A.$$
- on the left of ρ^*, the curve $\Upsilon_{K_{out}}$ is below Λ and is shifted less than $C^2\rho_m$ above. The possible intersection between $\Upsilon_{K'_{out}}$ and the line $f^* + df$ occurs thus on the right of ρ_1 with
$$|\rho_1 - \rho^*| = \frac{df + C^2 A\rho_m}{a} \leq \frac{C^1 + C^2\rho_m}{a} A.$$

The new state on the outgoing road being between the lines $\rho v = f^* - |df|$ and $\rho v = f^* + |df|$ and on the curve $\Upsilon_{K'_{out}}$, the new density $\bar{\rho}$ is between ρ_1 and ρ_2. This leads to
$$|\rho^* - \bar{\rho}| \leq \frac{C^1 + C^2\rho_m}{a} A = C^3 A.$$

4. The total variation on the outgoing road can thus be bounded
$$\begin{aligned}
|\rho^* - \bar{\rho}| &\leq C^3 A \\
|v^* - \bar{v}| &\leq \frac{f^* + df}{\rho^* - C^3 A} - \frac{f^*}{\rho^*} \\
&\leq \frac{f^* + C^1 A}{\rho^* - C^3 A} - \frac{f^*}{\rho^*} = \frac{C^1 A}{\rho^* - C^3 A} + \left(\frac{f^*}{\rho^* - C^3 A} - \frac{f^*}{\rho^*}\right).
\end{aligned} \tag{5.24}$$

Taking A small enough such that $C^3 A \leq \frac{\rho^*}{2}$,
$$\begin{aligned}
|v^* - \bar{v}| &\leq \frac{2C^1}{\rho^*} A + \frac{C^3 f^* A}{\rho^*(\rho^* - C^3 A)} \\
&\leq \frac{2C^1}{\rho^*} A + \frac{2C^3 f^*}{{\rho^*}^2} A.
\end{aligned}$$

It can be concluded that there exists a constant C such that
$$|\rho^* - \bar{\rho}| + |v^* - \bar{v}| \leq C(|d\rho| + |dv|)$$
and this equilibrium is stable.

Third claim

If all the states on the incoming roads belong to $\mathring{D}_2$ and at most one state on the outgoing roads belongs to Γ then the equilibrium is stable.

The proof is divided in two parts. First the influence of a perturbation occurring on an outgoing road is analysed and then of a perturbation on an incoming road.

Perturbation on an outgoing road

As explained in the first point of Section 5.3.5, only perturbations coming from an outgoing road which state belongs to Γ need to be analysed. If the perturbation occurs on such road, this perturbation may only be a shock wave since it is the only wave with a negative speed.

In Figure 5.11 are represented the initial state 1 and the new state after the perturbation (state 2). Nothing having changed on the incoming roads, the solution will belong to the same curve $\Upsilon_{K_{out}}$ and the total flow able to leave the incoming road is at least f^*. The final state on the outgoing road will be the state 3 which is the state on the curve $\Upsilon_{K_{out}}$ and in the new admissible region (function of the speed $v^* + dv$) with the maximal flow (see Figure 5.6(c)).

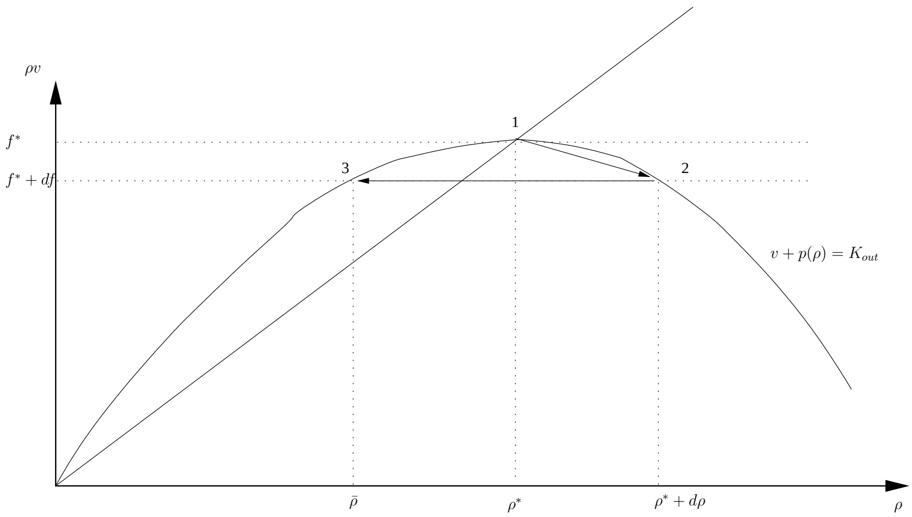


Figure 5.11: Bounds on the variation of ρ and f on the outgoing road. The state (ρ^*, f^*) is the original state before the perturbation.

For this analysis, the following notations are adopted

$$\begin{aligned} f(d\rho) &= K(\rho^* + d\rho) - (\rho^* + d\rho)p(\rho^* + d\rho) - f^* \\ d\bar{\rho} &= \bar{\rho} - \rho^* \end{aligned}$$

where $f \in C^\infty$.

The proof is done in three steps:

1. A bound on the final variation of ρ and v on the outgoing road can be obtained. This is done by obtaining first a bound on the flow, then on the density and finally on the speed.

For the flow, the final variation is equal to the initial variation which can be bounded in function of $d\rho$:

$$df \leq \|f'\|_\infty d\rho.$$

To find the bound on the density, a Taylor series expansion can be made around the origin of the function f:

$$f(\rho) = -a\rho^2 + \frac{f'''(\rho')}{3!}\rho^3 \qquad \text{with } 0 \leq \rho' \leq \rho,\ a > 0.$$

The goal is to show that

$$\exists \epsilon, C' > 0 \quad \text{such as} \quad \forall |d\rho| \leq \epsilon,\ \ |d\bar{\rho}| \leq C' |d\rho|.$$

In the worst case of the Taylor development, i.e. when the smallest perturbation of $d\rho$ produces the biggest change $d\bar{\rho}$, one has

$$f(\rho) = -a\rho^2 - \frac{\|f'''\|_\infty}{3!}\rho^3.$$

In this case, a perturbation $d\rho$ implies

$$df = -ad\rho^2 - \frac{\|f'''\|_\infty}{3!}d\rho^3.$$

With any $C' > 1$ and $d\rho$ small enough,

$$f(-C'd\rho) = -a(-C'd\rho)^2 - \frac{\|f'''\|_\infty}{3!}(-C'd\rho)^3 < df \tag{5.25}$$

thus

$$|d\bar{\rho}| \leq C'd\rho$$

since $d\bar{\rho}$ is defined such that $f(d\bar{\rho}) = f(d\rho)$. To prove (5.25), simply check that

$$-a(-C'd\rho)^2 - \frac{\|f'''\|_\infty}{3!}(-C'd\rho)^3 < -ad\rho^2 - \frac{\|f'''\|_\infty}{3!}d\rho^3$$

is equivalent to

$$a(C'^2 - 1)d\rho^2 > \frac{\|f'''\|_\infty}{3!}(1 + C'^3)d\rho^3$$

which is true for $d\rho$ small enough since $cx^2 > x^3 \quad \forall x \in [0, c)$.

On the basis of the $|d\rho|$ and $|dv|$ bounds, the final variation of speed on the outgoing road can be bounded using the same reasoning as in (5.24).

It can be concluded that, on the outgoing road, there exists a constant C such that

$$|\rho^* - \bar{\rho}| + |v^* - \bar{v}| \leq C(|d\rho| + |dv|)$$

for $|d\rho|$ and $|dv|$ small enough.

2. Due to the variation of the admissible region on the outgoing road, the states on the incoming roads may also have changed. The final variation on one of the incoming roads can be bounded as follow:

   - On the incoming roads, there is a decrease of flow less than df.
   - In term of density, one has (see Fig. 5.12)

   $$|\bar{\rho} - \rho^*| \leq \frac{1}{a} df \tag{5.26}$$

   where $a > 0$ is the slope of the line passing through (ρ^*, f^*) and $(\sigma_K, \gamma_K(\sigma_K))$.
   - Since $\left|\bar{f} - f^*\right|$ and $|\bar{\rho} - \rho^*|$ are bounded in function of $d\rho$ and dv, by the same reasoning as in (5.24), this is also true for $|\bar{v} - v^*|$.

   It can be concluded that there exists a constant C such that

   $$|\rho^* - \bar{\rho}| + |v^* - \bar{v}| \leq C(|d\rho| + |dv|)$$

   on the incoming road.

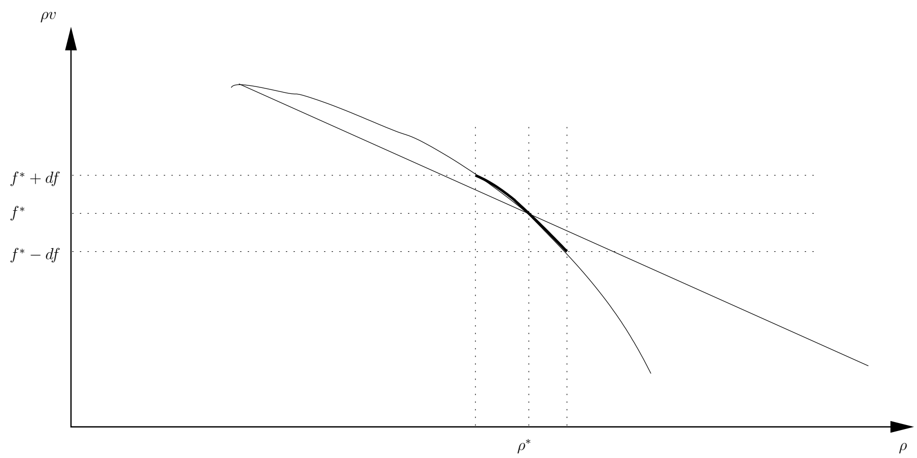


Figure 5.12: Maximum variation of the density for a fixed maximal variation of the flow on an incoming road.

3. Eventually this decrease of the flow leaving the incoming road may influence the state on a second outgoing road (which must belong by assumption to $\mathring{D}_1$). This variation can be bounded as in Section 5.3.5.

Perturbation on an incoming road
If the perturbation occurs on the incoming road and is small enough, this perturbation may only be a contact discontinuity since this is the only wave

with a positive speed. This case is represented in Figure 5.13. This perturbation will influence the states of the outgoing roads. Since the variation on the outgoing road belonging to $\mathring{D}_1$ can be bounded like in Section 5.3.5, it is only necessary to consider the case where the outgoing road belongs to Γ.

Two cases must be considered: when the density on the incoming road decreases and when it increases.

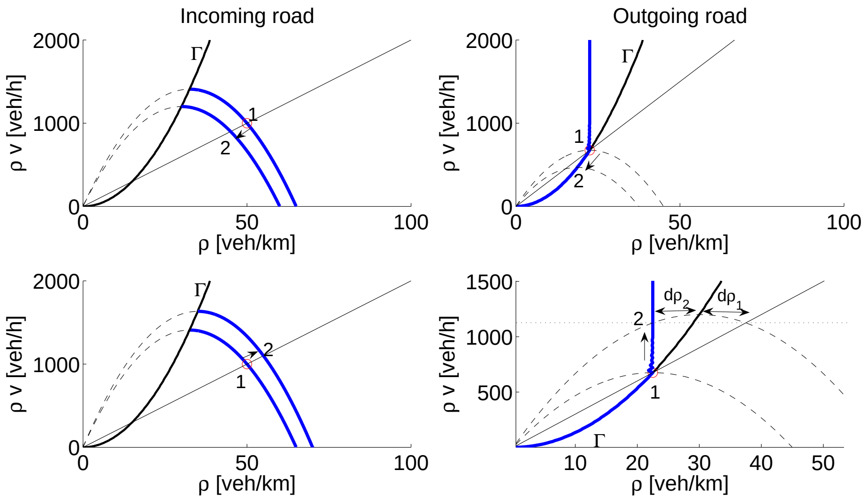


Figure 5.13: The two top figures correspond to a decrease of density on the incoming road while the two bottom ones correspond to an increase of density.

When the density on the incoming road decreases, the new state on this incoming road is on a curve $\Upsilon_{K'_{in}}$ with $K'_{in} \leq K_{in}$. This decrease of K_{in} can easily be bounded in function of $d\rho$:

$$|dK_{in}| \leq \|p'\|_\infty |d\rho|. \tag{5.27}$$

On the outgoing road, the state will belong to a new curve $\Upsilon_{K'_{out}}$ with

$$|dK_{out}| \leq |dK_{in}|. \tag{5.28}$$

Since too much flow is trying to pass from the incoming roads to the outgoing one (this was already the case before the perturbation), the new state on the outgoing road will be the maximum of the curve $\Upsilon_{K'_{out}}$. To prove that the variation of the traffic variable ρ and v on the outgoing road can be bounded in function on $d\rho$, from (5.27) and (5.28), it is sufficient to prove that the variation of $(\sigma_K, \gamma_K(\sigma_K))$ may be bounded by the variation of dK.

Let $g(\rho)$ be defined as $((\rho p(\rho))')^{-1}$ then $g' = \frac{1}{(\rho p(\rho))''} \in C^0$ and thus $g \in C^1$. It is easy to see that

$$\sigma_K = g(K)$$

and thus

$$|d\sigma_K| \leq \|g'\|_\infty |dK|.$$

It can also easily be computed that

$$\gamma_K(\sigma_K) = Kg(K) - g(K)p(g(K)) = h(K).$$

Since g and p are C^1, $h(K)$ is C^1 and

$$d\gamma_K(\sigma_K) \leq \|h'\|_\infty dK.$$

Having a bound on the variations of the flow and the density in function of $d\rho$, a bound on the variation of the speed can be deduced. It can thus be concluded that there exists C such that

$$|\rho^* - \bar{\rho}| + |v^* - \bar{v}| \leq C(|d\rho| + |dv|)$$

on the outgoing road.

On the incoming road, it is known that the final state is on the curve $\Upsilon_{K'_{in}}$. Moreover the variation of flow is linked to the variation of flow on the outgoing road and thus can be bounded in function of $d\rho$. Using the same reasoning as in (5.26), one can bound the total variation of the density and speed on the incoming road in function of $d\rho$.

When the density on the incoming road increases, this corresponds to the situation of the two bottom figures in Fig. 5.13. The proof is a combination of the previous explanations. On the outgoing road:

- the new state belongs to the curve $\Upsilon_{K'_{out}}$ with dK_{out} bounded in function of $d\rho$;
- the state with the maximum flow belonging to $\Upsilon_{K'_{out}}$ inside the admissible region is the state 2 of Figure 5.13;
- since dK_{out} is bounded, the difference of the flow df' associated to the maxima of $\Upsilon_{K_{out}}$ and $\Upsilon_{K'_{out}}$ is bounded;
- the variation of flow between state 2 and state 1 is bounded by df';
- for the density, $d\rho_1$ is bounded by df'/v^*;
- $d\rho_2$ may be bounded in function of $d\rho_1$;
- the final variation of density on the outgoing road is less than $d\rho_2$;
- since the variation of flow and density may be bounded, the variation of speed may also be bounded.

On the incoming road, since the variation of flow is bounded, this is also true for the density and the speed.

Since for all small variations and on all roads there exists C such that

$$|\rho^* - \bar{\rho}| + |v^* - \bar{v}| \leq C(|d\rho| + |dv|)$$

then the equilibrium is stable.

5.3.6 Existence of a solution in the BV space

In the previous sections, only solutions to Riemann problems at the junction were considered. Using the previous stability results, it is possible to prove the existence of a solution for a larger class of initial conditions (see [20]).

If one considers a network with only one junction and an initial state $((\rho_{1,0}(x), v_{1,0}(x)),\ (\rho_{2,0}(x), v_{2,0}(x)),\ (\rho_{3,0}(x), v_{3,0}(x)))$ which is a stable equilibrium such as

$$\|(\rho_{i,0}(x), v_{i,0}(x))\|_{BV} \leq \epsilon$$

and

$$\sup_x |\rho_{i,0}(x) - \rho_{i,0}| + \sup_x |v_{i,0}(x) - v_{i,0}| \leq \epsilon,$$

then for ϵ small enough, there exists a weak solution satisfying the initial conditions and the equilibrium conditions at the junction.

In this equation BV denotes the space of functions with bounded variation. The corresponding norm $\| \ \|_{BV}$ is defined by

$$\|u\|_{BV} = \sup\{v(u, P) : P \text{ is a partition of } [a, b]\}$$

where

$$v(u, P) = \sum_{k=1}^{n} d(u(t_k), u(t_{k-1}))$$

with $P = \{x_0 < t_1 < \cdots < x_n\}$ a partition of the domain of u.

Proof. For the complete proof see [20] and [6]. □

5.3.7 Comparison with other models

To the author's knowledge, the only other junction models developed on the basis of the AR single road model are presented in [32, 33], [20] and [50]. As already explained, the model [50] is very close to our model except that the sharing coefficients for a merge junction are fixed in [50] and that the particular function $p(\rho) = -V(\rho)$ is used.

The idea developed in [32] is similar to the idea presented in the present work: each driver tries to conserve his behaviour and the resulting behaviour on the outgoing road is a mean of the different incoming behaviours. The main difference is that [32] analyses the mixing problem on the basis of an underlying microscopic model whereas our approach stays at a macroscopic scale with the formulation

$$v_a + p(\rho_a) = \sum_i \alpha_i (v_i + p(\rho_i)) \tag{5.29}$$

The approach of [32] leads to a homogenisation problem which may need to introduce a new function $p(\rho)$ after the junction. The model presented in the present work can therefore be seen as an approximation of this homogenisation problem. The main advantage of (5.29) is its simpler formulation leading to easier computations.

It is worth mentioning that the two approaches are close and even identical for a diverging junction. Indeed, in this case, the two models predict that the quantities $v + p(\rho)$ on the outgoing roads are the same as the quantity $v + p(\rho)$ on the incoming road. Since the two models select the new values of (ρ_i, v_i) near the junction by performing a maximisation of the total passing flow, they lead to the same values. For a merge junction, the situation is more complex since [32] may introduce a new function $p(\rho)$ after the junction.

There are some important differences between the model presented in this work and the model of [20]:

1. In our model the existence of the invariant region $\mathcal{R}$ is a natural consequence of the choice made for the additional criterion ("conservation" of $v + p(\rho)$). In the model [20], the same invariant region exists but is a consequence of the introduction of a maximal speed in the description of the admissible region of an outgoing road.

2. Our model is able to represent the capacity drop phenomenon while this is not the case for the model [20]. The main reason is the choice made for the order in which some variables are fixed or maximised on the outgoing road:

   - In our model, the order is
     (a) fix the admissible region
     (b) fix the $v + p(\rho)$ value
     (c) maximise the flow
   - In the model [20], the order is
     (a) fix the admissible region
     (b) maximise the flow and then take it as a constraint
     (c) maximise the speed/density or minimise the total variation of ρ along the solution on the outgoing roads.

   In our model, the maximisation of the flow is the final criterion used to choose the solution while in the model [20] this maximisation occurs before any other additional criterion.

3. One limitation of our model is that it is developed only for 2-incoming-1-outgoing and 1-incoming-2-outgoing roads junctions. The model [20] is developed for n-incoming-m-outgoing roads junction but also has an important limitation, i.e. $m \geq n$ which excludes the most common merge junction.

4. The two models have more or less the same properties when dealing with stability of equilibrium and existence of a solution around a stable equilibrium.

Chapter 6

Conclusions

Two aspects of the traffic problems were addressed in this thesis: the modelling and the control. The modelling part consists in deriving equations which allow to predict the evolution of the traffic state. The control part consists in designing rules which improve the global efficiency of the road network.

Modelling aspects

Our main interest being the control of the traffic, the predictions of our models must be sufficiently accurate for the density and the speed of the vehicles. Indeed, these two values are used to compute the efficiency of the road network. Based on such modelling, different control strategies can be developed and evaluated.

Two network models are developed in this work. The first one is an existing model based on the LWR single road model that we have extended to take into account the capacity drop phenomenon. The second one is an original model based on the AR single road model. Each of these models are able to represent the capacity drop phenomenon. This property is important since this phenomenon can produce an hysteresis effect which can have significant influence when considering the global network efficiency. None of these models present incoherences such as negative speed, infinite density, drivers acting in function of what happens behind them and so forth. The first model is simpler, being able to represent less phenomena than the second model, but very easy to implement; it allows fast computation. Depending on the precision needed and the computing power available, one will choose to use the simple LWR model or the more precise AR model.

This second road network model has been developed by adding only one new assumption: "conservation" of the quantity $v + p(\rho)$ representing the behaviour of the drivers; the other assumptions are the commonly admitted ones: conservation of the flow, sharing of the available space on the outgoing road based on coefficients, drivers acting in order to maximise the passing flow. This model is coherent, realistic and able to represent the capacity drop phenomenon. Our junction model doesn't add any new parameters compared to those introduced by the AR single road model (the function $p(\rho)$). Combined with this single road model, it provides a complete description of the traffic evolution on a road network.

The extension of our junction model to n-incoming-1-outgoing and 1-incoming-n-outgoing roads is straightforward. It does not need to introduce additional assumptions. If multiple incoming and multiple outgoing roads are considered, the extension is more difficult: additional criteria must be added in order to describe how the different flows interact.

The next important work in the modelling part is the validation of our two developed models using experimental measurements. A first validation has been done for the single order LWR model (see Section 3.3.5) which has led to satisfying results. However, due to the lack of data on the on/off–ramps, these data were reconstructed using the available data. This process leads to smaller confidence in the validation accuracy. For the AR model, such validation has not yet been performed.

Control aspects

The section of this thesis dealing with the control aspects can mainly be divided in three parts: static optimisations, a routing strategy and a linear feedback ramp metering method.

In the first part, a realistic cost function has been developed to represent the global efficiency of a road network. This function includes the costs related to the spent time, the generated pollution and the risk of accident. On the basis of this cost function various optimisations can be performed to analyse the best flow repartitions, the influence of some speed limitations or the construction of a new road. It turns out that the best solution depends usually on the relative importance of the time cost and the pollution cost which is a political choice.

In the second part, a routing strategy has been developed. This control method expresses which fraction of the inflow must be redirected to an alternative path. The idea behind this strategy is to use a Proportional–Integral controller which minimises an approximation of the total time spent. This control method is easy to implement and, using numerical simulations, has been shown to allow a 15 % improvement of the global cost.

In the third part, a ramp metering strategy using linear feedback has been developed. This control method has been demonstrated to ensure stability with respect to small perturbations. Another benefit of this strategy is its implementation easiness. Indeed only sampled measurements are needed. This was not the case of the previous routing strategy where the complete traffic states (or an approximated reconstruction) on the two paths were needed. The stability problem involved in this ramp metering strategy has been shown to induce a spectral radius problem. Based on this matrix problem, theoretical results and an algorithm describing the maximal values of the linear feedback coefficients which still ensure stability are derived.

The two previous control methods have been developed and illustrated using the LWR model but can also be used with the second order model. The main element of the routing strategy is the function f which represents the flow associated to the density. If the AR second order model is used with the presence of a relaxation term $\frac{V(\rho)-v}{\tau}$, the function f can be replaced by $\rho V(\rho)$. If the relaxation coefficient τ is small enough, then the routing strategy should lead to more or less the same results.

The second control method, the linear feedback ramp metering, can also be used with the second order AR model. The only mathematical difficulty is to express the linearisation of the boundary conditions and the control inputs. In practice a major difficulty remains: the feedback strategy provides values for the density and the speed that must be applied at the on-ramps. The network operator should thus be able to impose simultaneously the density and the speed. This is not easy and cannot certainly be done by just using a simple traffic light.

Some work can still be done in order to extend the two previous control strategies. For the routing strategy, it could be interesting to consider the possible coordination problems. Indeed, if there are different pairs of paths with some overlaps then it might be possible than the coefficients used for the Proportional–Integral controller have an influence on the global stability.

Another interesting question for the second strategy, the linear feedback ramp metering, is how to choose the values of the control coefficients on the basis of some transfer function. The main problem with this question is that the usual ways (minimisation of the 2-norm or the ∞-norm) cannot be used. A possible solution is to consider weighted norms to ignore high frequency perturbations. This should prevent the particular perturbation represented in Fig. 4.22 to be considered and thus allow to select values for the coefficients. This should also make the 2-norm (4.29) finite.

Perspectives

The traffic problem is far from being solved. A lot of things have not been considered in this thesis.

The two models developed here are only valid for highways. One must still consider models for urban roads. Such models exist in the literature. In practice, when a large network with both highways and urban area is considered, the two types of model must be used and must be connected together. To the author's knowledge there is no work in the literature on the connection of a macroscopic model like the AR model and a microscopic model more suitable for urban area.

In this thesis the control strategies have been developed and illustrated using small networks with well defined boundaries. In practice the networks are interconnected and it is not obvious that two control strategies, independently well designed, will not badly interact. In practice there is no boundaries. When a feedback strategy prevents some drivers from entering a highway, a queue is growing somewhere on the network and can lead to a traffic jam.

The problem of the traffic control is also very complex and there is still no miraculous solution. However, the traffic problem is a subject with increasing popularity among the research community. This is probably because the traffic problem is becoming a major difficulty is some regions and that politicians are investing to find solutions. This increasing popularity also means that people from different communities are now working on the traffic problem. This can lead to duplicate work if the communications between these communities are not sufficient.

With increasing research capacities on the subject, it is likely that more and more useful results will be produced. Even if none of these results will be able to solve completely the traffic problem, one can hope that some will be able to reduce it.

Bibliography

[1] A. Aw, A. Klar, T. Materne, and M. Rascle. Derivation of continuum traffic flow models from microscopic follow-the-leader models. *SIAM J. APPL. MATH.*, 63:259–278, 2002.

[2] A. Aw and M. Rascle. Ressurection of "second order" models of traffic flow. *SIAM J. APPL. MATH.*, 60:916–938, 2000.

[3] C. Bardos, A. LeRoux, and J. Nedelec. First order quasilinear equations with boundary conditions. *Communications in Partial Differential Equations*, 1979.

[4] T. Bass. Road to ruin. *Discover*, pages 56–61, May 1992.

[5] A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences.* siam, 1994.

[6] Alberto Bressan. *Hyperbolic Systems of Conservation Laws – The One-dimensional Cauchy Problem.* Oxford University Press, 2000.

[7] M. J. Cassidy and R. L. Bertini. Some traffic features at freeway bottlenecks. *Transportation Research Part B*, B33:25–42, 1999.

[8] G. M. Coclite, M. Garavello, and B. Piccoli. Traffic flow on a road network. *SIAM J. Math. Anal.*, 36(6):1862–1886, 2005.

[9] E. D. Conway and J. A. Smoller. Shocks violating Lax's condition are unstable. In *Proceedings of the american mathematical society*, volume 39, pages 353–356, July 1973.

[10] Carlos F. Daganzo. Requiem for second-order fluid approximations of traffic flow. *Transportation Research B*, 4:277–286, 1995.

[11] C.F. Daganzo. The cell transmission model 2: network simulation. *Transportation Research B*, 29B(2):79–93, 1995.

[12] C.F. Daganzo. A finite difference approximation of the kinematic wave model of traffic flow. *Transportation Research B*, 29:261–276, 1995.

[13] Ortuzar Juan de Dios and Willumsen Luis G. *Modelling transport, 2nd edition.* Wiley, 1994.

[14] J. de Halleux, C. Prieur, J.-M. Coron, B. d'Andréa Novel, and G. Bastin. Boundary feedback control in networks of open channels. *Automatica*, 39:1365–1376, 2003.

[15] John C. Doyle, Bruce A. Francis, and Allen R. Tannenbaum. *Feedback control theory*. MacMillan, New York, 1992.

[16] F. Dunkerley and S. Proost. Note on sustainable cost function. Internal report from the ETE unit of the K.U.Leuven: Sustainable effects of traffic management systems, 2004.

[17] F. R. Gantmacher. *The Theory of Matrices*. New York, NY, USA, 1959.

[18] Song Gao and Ismail Chabini. Optimal routing policy problem in sotchastic time-dependent networks. *Transportation Research B*, 40:93–122, 2006.

[19] M. Garavello and B. Piccoli. *Traffic Flow on Networks*. 2006.

[20] Mauro Garavello and Benedetto Piccoli. Traffic flow on a road network using the Aw–Rascle model. *Comm. Partial Differential Equations*, 31(2):243–275, 2006.

[21] D.C. Gazis, R. Herman, and R. Rothery. Nonlinear follow-the-leader models of traffic flow. *Oper. Res.*, (9):545–567, 1961.

[22] F. L. Hall and K. Agyemang-Duah. Freeway capacity drop and the definition of capacity. *Transportation Record*, (1320):99–109, 1991.

[23] B. Haut and G. Bastin. A second order model for road traffic networks. In *Proceedings of ITSC'05 – 8th International IEEE Conference on Intelligent Transportation Systems*, pages 178–184, september 2005.

[24] B. Haut and G. Bastin. A system-optimal routing policy for road traffic networks. In *Proceedings of 11th IFAC Symposium on Control in Transportation Systems*, Delft, The Netherland, August 29-30-31 2006.

[25] B. Haut and G. Bastin. A second order model of road junctions in fluid models of traffic networks. *AIMS Journal on Networks and Heterogeneous Media (NHM)*, 2(2):227–253, 2007.

[26] B. Haut, G. Bastin, and Y. Chitour. A macroscopic traffic model for road networks with a representation of the capacity drop phenomenon at the junction. In *Proceedings of 16th IFAC World Congress*, Prague, Czech Republic, July 4–8 2005.

[27] B. Haut, G. Bastin, and P. Van Dooren. Maximal stability region of a perturbed nonnegative matrix. *submitted for publication.*

[28] B. Haut, G. Bastin, and P. Van Dooren. Maximal nonnegative perturbation of a nonnegative matrix. In *Positive Systems*, volume 341 of *Lecture Notes in Control and Information Sciences*, pages 201–208. Springer, 2006.

[29] B. Haut, H. Cote, G. Bastin, and G. Campion. DWTC-CP/40 Final report: Sustainability effects of traffic management systems, 2005.

[30] Hawkins. *The Ecology of Commerce: A Declaration of Sustainability*. Harper Collins Publishers, NY, USA, 1993.

[31] M. Herty and A. Klar. Modeling, simulation, and optimization of traffic flow networks. *SIAM J. Sci. Comput.*, 25(3):1066–1087, 2003.

[32] M. Herty, S. Moutari, and M. Rascle. Optimization criteria for modelling intersections of vehicular traffic flow. *Networks and Heterogeneous Media*, 1(2):275–294, June 2006.

[33] M. Herty and M. Rascle. Coupling conditions for a class of second-order models for traffic flow. *SIAM Journal on Mathematical Analysis*, 38(2):595–616, 2006.

[34] D. Hinrichsen and N. K. Son. Stability radii of positive discrete-time systems under affine parameter perturbations. *International J. Robust and Nonlinear Control*, 8:1169–1188, 1998.

[35] H. Holden and N. H. Risebro. A mathematical model of traffic flow on a network of unidirectional roads. *Siam J. Math. Anal.*, 26(4):999–1017, 1995.

[36] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[37] Hai-Jun Huang and William H.K. Lam. Modeling and solving the dynamic user equilibrium route and departure time choice problem in network with queues. *Transportation Research B*, 36:253–273, 2002.

[38] W.L. Jin and H.M. Zhang. On the distribution schemes for determining flows through a merge. *Transportation Research PartB*, (37):521–540, 2003.

[39] G. Kolata. What if they closed 42nd street and nobody noticed? *The New York Times*, December 25 1990.

[40] R.D. Kühne and R. Beckschulte. Non-linearity stochastics in unstable traffic flow. *Transportation and traffic theory*, pages 367–386, 1993. Proc 12th Int. Symp on Trans. and Traffic Theory, Berkeley, CA. New York; Elsevier.

[41] J.-P. Lebacque, S. Mammar, and H. Haj Salem. Generic second order traffic flow modelling. In *Proc. of the 17th ISTTT (International Symposium on Transporation and Traffic Theory)*, London, 2007.

[42] J.P Lebacque. Semimacroscopic simulation of urban traffic. In *Int. 84 Minneapolis Summer Conference. AMSE.*, 1984.

[43] J.P. Lebacque. The Godunov scheme and what it means for first order traffic flow models. In *Transportation and traffic flow theory, Proceedings of the 13th ISTTT,* Pergamon, 1996.

[44] J.P. Lebacque. Intersection modeling, application to macroscopic network traffic flow models and traffic management. In S. Hoogendoorn and M. Schreckenberg, editors, *Proceedings of Traffic and Granular Flow'03.* Springer, 2004.

[45] J.P. Lebacque and H. Haj-Salem. Ramp metering, speed management and braess-like paradoxes. TRISTAN V, 2004.

[46] J.P. Lebacque and M.M. Khoshyaran. Modelling vehicular traffic flow on networks using macroscopic models. In *Proc. of the FVCA II*, pages 551–558, 1999.

[47] J.P. Lebacque and M.M. Khoshyaran. First order macroscopic traffic flow models: intersection modeling, network modeling. In *Proc of the 16th ISTTT Conf*, pages 365–386, 2005.

[48] Ta-Tsien Li. *Global classical solutions for quasi-linear hyperbolic systems.* Research in Applied Mathematics, Wiley, 1994.

[49] M. J. Lighthill and J. B. Whitham. On kinematic waves. i: Flow movement in long rivers. ii: A theory of traffic flow on long crowded roads. *Proc. Royal Soc. London Ser. A*, (229):281–345, 1955.

[50] S. Mammar, J.-P. Lebacque, and H. Haj Salem. Intersection modelling based on the second order traffic flow model (arz). In *Proc. of the 11th Euro Working Group on Transportation*, pages 570–580, Bari, 2005.

[51] M. Papageorgiou, H. Hadj-Salem, and J.-M. Blosseville. Alinea: a local feedback control law for on-ramp metering. *Transportation Research Record*, (1320):58–64, 1991.

[52] Markos Papageorgiou. Freeway ramp metering: An overview. *IEEE transactions on intelligent transportation systems*, 3(4):271–281, december 2002.

[53] E.I. Pas and S.L. Principio. Braess' paradox: Some new insight. *Transportation Research B*, 31:265–276, 1997.

[54] H.J. Payne. Models of freeway traffic and control. *Simulation Council Proceedings*, 1:51–60, 1971.

[55] I. Prigogine and F.C. Andrews. A Boltzmann like approach for traffic flow. *Oper. Res.*, (8):789–797, 1960.

[56] D. Serre. *Systèmes de lois de conservation.* Diderot, Art et Sciences, 1996.

[57] C.W. Shu. Essentially non-oscillatory and wieghted essentially non-oscillatory schemes for hyperbolic conservation laws. In *Advanced Numerical Approximation of Nonlinear Hyperbolic Equations*, volume 1697 of *Lectures Notes in Mathematics*, pages 325–432. Springer, New York, 1998.

[58] R. Tarjan. Depth first search and linear graph algorithms. *SIAM Journal of Computing*, 1(2):140–160, june 1972.

[59] Ybing Wang, Goerge Sarros, Markos Papageorgiou, and Willem Jan Knibbe. Real-time route guidance in large-scale urban express ring-roads. In *Proceedings of the 16th IFAC World Congress*, July 2005.

[60] G.B. Whitham. *Linear and nonlinear waves.* 1973.

[61] H.M. Zhang. A non-equilibrum traffic model devoid of gas-like behavior. *Transportation Research B*, 36:275–290, 2002.