# The dynamic user equilibrium on a transport network: mathematical properties and economic applications

Nicolas Wagner

**UNIVERSITÉ PARIS-EST**

École doctorale Ville Environment Territoire
Laboratoire Ville Mobilité Transport

**THÈSE DE DOCTORAT**

Spécialité Économie des transports

# The dynamic user equilibrium on a transport network: mathematical properties and economic applications

**p**ar Nicolas Wagner

Directeur de thèse : Fabien Leurent
Co-directeur de thèse : Vincent Aguiléra

Présentée et soutenue publiquement le 24 janvier 2012

JURY

| | | |
|---|---|---|
| Erik Verhoef | Vrije Universiteit Amsterdam | *Président* |
| Guido Gentile | Sapienza Università di Roma | *Rapporteur* |
| Jèrôme Renault | Toulouse School Of Economics | *Rapporteur* |
| Jean Delons | Cofiroute - Dpt Economie Trafic | *Examinateur* |
| Jean-Patrick Lebacque | IFSTTAR | *Examinateur* |
| Fabien Leurent | Université Paris-Est | *Directeur* |

# Abstract Summary

This thesis is focused on dynamic user equilibrium models and their applications to traffic assignment. It aims at providing a mathematically rigorous and general formulation for the dynamic user equilibrium. Particular attention is paid to the representation of transport demand and more specifically to trip scheduling and users with heterogeneous preferences. This is achieved by expressing the dynamic user equilibrium as a Nash game with a continuum of players. This allows for a precise, concise and microeconomically consistent description.

This thesis also deals with computational techniques. We solve analytically equilibrium on small networks to get a general intuition of the complex linkage between the demand and supply of transport in dynamic frameworks. The intuition acquired from the resolution is used to elaborate efficient numerical solving methods that can be applied to large size, real life, transport networks.

Along the thesis several economic applications are proposed. All of them are dealing with the assessment of congestion pricing policies where are likely to reschedule their trips. In particular, a pricing scheme designed to ease congestion during holiday departure periods is tested. In this scheme a toll varying *within the day* and *from day to day* is set on the french motorway network. This form to toll is especially appealing as it enables the operator to influence the departure day as well as the departure time. Indeed it is shown that even moderate variations of the toll with time might have strong impacts on an highly congested interurban network.

# Résumé Court

Cette thèse porte sur les modèles d'équilibres dynamiques sur un réseau de transport et leurs applications à l'affectation de trafic. Elle tente d'en propose une formulation à la fois générale et mathématiquement rigoureuse. Une attention particulière est accordée à la représentation de la demande de transport. Plus spécifiquement, la modélisation de l'hétérogénéité dans les préférences des usagers d'un réseau de transport, ainsi que des stratégies de choix d'horaire dans les déplacements, occupe une place importante dans notre approche. Une caractéristique de ce travail est son fort recours au formalisme mathématique; cela nous permet d'obtenir une formulation concise et micro-économiquement cohérente des réseaux de transport et de la demande de transport dans un contexte dynamique.

Cette thèse traite aussi de méthodes de résolution en lien avec les modèles d'équilibres dynamiques. Nous établissons analytiquement des équilibres sur des réseaux de petites tailles afin d'améliorer la connaissance qualitative de l'interaction entre offre et demande dans ce contexte. L'intuition retirée de ces exercices nous permet de concevoir des méthodes numériques de calculs qui peuvent être appliquées à des réseaux de transport de grande taille.

Tout au long de la thèse plusieurs applications économiques de ces travaux sont explorées. Toutes traitent des politiques de tarification de la congestion et de leurs évaluation, notamment lorsque les automobilistes sont susceptibles d'ajuster leurs horaires de départ. En particulier une politique tarifaire conçue pour limiter la congestion lors des grands départs de vacances est testée. Elle consiste à mettre en place un péage sur le réseau autoroutier variant selon l'heure de la journée mais aussi de jour en jour. Ce type de péage est particulièrement intéressant pour les exploitants car il leur permet d'influencer à la fois sur l'heure et le jour de départ des vacanciers. Les méthodes développées dans cette thèse permettent d'établir que les gains en termes de réduction de la congestion sont substantiels.

# Contents

# List of Figures

# List of Tables

# Abstract

This thesis is focused on dynamic user equilibrium models for traffic assignment. It aims at providing a mathematically rigorous and general formulation for the dynamic user equilibrium. Particular attention is paid to the representation of transport demand and more specifically to trip scheduling and users with heterogeneous preferences. This work is characterized by a high level of mathematical formalism; this allows for a precise, concise and microeconomically consistent description of dynamic transport networks and dynamic transport demand.

Although the rigorous formalization of dynamic user equilibrium models is the main object of the thesis, it also deals with computational techniques. We aim at solving analytically some stylized models to get a general intuition of the complex linkage between the demand and supply of transport in dynamic frameworks. The intuition acquired from solving these analytical models will be used to elaborate efficient numerical solving methods that can be applied to large size, real life, transport networks.

Our approach can be broken down into four steps. The first step is the literature review (Part I). Extensive reviews of academic works constitute the first stage of this study. Second, a game theoretic formulation of the dynamic user equilibrium is proposed (Part II). It strongly relies on up-to-date results from mathematical economics on games with a continuum of players. Third, analytical resolutions of this model are presented in restricted cases (Part III). Although these specific cases are chosen to answer specific issues in transport economics, they gave us interesting insights regarding the mathematical structure of the problem. In particular they have been very valuable for the last step of this thesis (Part IV), where a computable model is designed and corresponding solution methods are proposed.

## Part I: Bibliography

Our literature review aims at formulating existing dynamic user equilibrium (DUE) models with a unified set of notations. This is a natural first step in our quest for a general framework for DUE models.

Chapter 1 is entitled *Dynamic network modelling and algorithmics*. DUE models operate on transport networks that are both time-varying and prone to congestion. It is thus essential to first precisely defined the network model that is used. This leads us to formalize a model of dynamic transport networks (DTN). In a DTN, arc travel times and costs are time-varying and so are the flows of traffic. Using this formalism, two algorithmic problems are presented. The first one, known as the continuous dynamic network loading problem, consists in determining the traffic flow propagation in a DTN and to deduce the resulting arc travel times and costs. The second one is the time-varying shortest route problem. In both cases, the DTN model allows to present and compare existing numerical schemes from the literature.

Chapter 2, *Mathematical formulations for the dynamic user equilibrium*, is divided in two parts. First, the most common DUE model, which we will simply referred to as the dynamic Wardrop equilibrium, is described. In this dynamic Wardrop assignment problem, users are homogeneous and might only choose their route. This specific DUE is reviewed in depth, as the literature covering it is vast and extremely rich. A particular attention is paid to the equivalence between each formulations. The associated algorithmic problem, the well-known *dynamic traffic assignment* problem, is also reviewed and the most common algorithms are presented and compared.

Then, extensions of the dynamic user equilibriums that considers more complex representations of the demand are considered. In particular, models including trip-scheduling and user heterogeneity are presented.

This bibliographic review shows that the literature leaves a number of questions unanswered, especially regarding the dynamic representation of the transport demand. In particular:

- *The mathematical properties of user equilibriums remain to be fully established*: existence results for dynamic equilibrium models have been shown only in specific cases and no uniqueness and stability results have been proven (Mounce, 2007). Along the same lines, mathematically concise formulations are rare, mainly due to the complexity of analytically formulating the traffic flow on a network.

- *User heterogeneity* is not fully represented: most of the existing analytical models consider a finite number of homogeneous groups of travellers, each group being characterized by a few variables *e.g.* vehicle type, value of time or preferred arrival time (for instance in De Palma and Marchal, 2002). A more general representation would be to consider continuous distributions over the space of characteristics. Although microeconomists have long considered small transport models with continuous heterogeneity (Vickrey's bottleneck model is probably the most famous example), up to now no general theoretical formulation is available.

- *More efficient algorithms for user equilibriums with departure time choice* are still required. User equilibriums with route choice can now be computed with reasonable efficiency on large size networks (Aguiléra and Leurent, 2009). A substantial amount of work is still needed to properly state the appropriate numerical solution techniques when the problem includes departure time choice. From an algorithmic viewpoint this is probably one of the most challenging problems in transport science currently.

## Part II: Dynamic congestion games and their application to dynamic traffic assignment

This part aims at designing a new framework for DUE models that allows a refined representation for the transport demand. A particular attention is paid to the representation of users heterogeneity and trip scheduling. To do so, recent results and models from mathematical economy and game theory are exploited.

Chapter 3, *Dynamic congestion games: presentation and a simple illustration*, presents a new category of games intended to be a new framework for dynamic user equilibrium models is introduced. These so-called dynamic congestion games offers a wide range of modelling possibilities. Users might be represented by a continuous distribution over one or many variables. Road pricing strategies can be embedded in the utility functions, possibly only for specific types of users and the pricing scheme might be time-varying. Finally, the possibility of intermediate stops from which the user might derive some utility, typically short shopping stops, might be taken in account. As far

as congestion modelling is concerned, the assumptions considered seems *a priori* weak and it is reasonable to think that they include a wide range of the operational models used for transport planning.

Two theoretical results are established within the chapter. First, a constructive proof of the existence and uniqueness of the solution to the dynamic traffic loading problem is exposed. Then, it is shown that the existence of a Nash equilibrium in dynamic congestion games is guaranteed under five natural assumptions on the congestion model.

Chapter 4, *Application to the dynamic traffic assignment problem on a network of bottlenecks*, establishes that the simplest dynamic user equilibrium model, known as the dynamic Wardrop equilibrium, can be seen as a particular case of dynamic congestion games. A known existence result, due to Mounce (2007), is then shown to derive from the existence result of Chapter 3.

## Part III: Analytical resolutions of simple games

Part III is devoted to two simple dynamic congestion games for which the solutions can be derived analytically. Both games are extension of Vickery's classical bottleneck model.

Chapter 5, *User equilibrium with general distribution of preferred arrival times*, studies the pattern of departure times at a single bottleneck, under general heterogeneous preferred arrival times. It generalizes Vickery's model, without the classical "S-shape" assumption *i.e.* that demand, represented by the flow rate of preferred arrival times, may only exceed bottleneck capacity on one peak interval. It delivers two main outputs. First, a generic analytical is given to solve the departure time choice equilibrium problem. Second, the graphical approach that pervades the solution scheme provides insights in the structure of the queued periods, especially so by characterizing the critical instants at which the entry flow switches from a loading rate (over capacity) to an unloading one (under capacity) and vice versa.

Chapter 6, *User equilibrium with continuously distributed values of time* presents a game with a two route network where users are continuously heterogeneous w.r.t. their value of time. Road infrastructures are assumed to present bottleneck congestion technology and a flat (*i.e.* time-invariant) toll is set on one of the routes while the other one is free. Using this framework, two pricing policies are assessed. In both cases, a toll is set on only one route

while the other is free. In the first policy the toll is set to maximize revenue, while in the second it is set to maximize the welfare gains. This known in the literature as a value pricing scheme and the two policies correspond respectively to the private and the public ownership of the tolled route. The analytical investigation demonstrates that the level of heterogeneity significantly impacts the efficiency (measured in terms of welfare gains) of each policy. The main result of this chapter is that the relative efficiency of the private ownership increases with the level of heterogeneity. This results is especially interesting to compare to the one of van den Berg and Verhoef (2010), that states the contrary when the toll is time-varying.

## Part IV: Numerical Resolution

Chapter 7, *A user equilibrium model with departure time choice*, presents a simple DUE model as a dynamic congestion game. The main features of the model are that time is represented continuously, that scheduling preferences are represented by continuous distributions of the preferred arrival times, that traffic flowing is multi-class and that time-varying tolls can be imposed on each arc. The main technical technical difference between this model and the ones currently developed in the literature is that the trip scheduling model is deterministic (contrary to Bellei, Gentile and Papola, 2005; De Palma and Marchal, 2002). It was designed to extend the LADTA model introduced by Leurent (2003*b*) whose original implementation did not account for departure time choice.

Now in a dynamic congestion games, users with the same characteristics may choose different departure times. This property is unconvenient from a computational perspective as it leads to memory-intensive representations of users' choices. Thus the possibility of imposing the same departure time for all users with the same characteristics is studied. In such a case the users' departure time distribution is said to be symmetric. It is shown that when users differ with their value of time and the network is subject to tolls, the existence of an equilibrium is no longer guaranteed. Hopefully, the users' arrival time distribution can be assumed to be symmetric without loosing the existence property. A restricted model of DUE, more suited to computation, is thus introduced. The subsequent chapters present algorithms to solve this restricted model.

Chapter 8, *A convex combination algorithm to compute the dynamic user*

*equilibrium*, is devoted to the presentation and assessment of an algorithm inspired from the classical convex combination scheme. The main algorithmic breakthrough is a numerical method that finds simultaneously the optimal arrival times of all users on an Origin-Destination pair. This method is shown to be significantly faster than the naïve approach where each user is treated separately. The whole algorithm is tested on small scale networks and the algorithm performs well on these examples. However it requires to set a parameter and this operation is time-consuming.

Chapter 9, *A user equilibrium computation algorithm based on user coordination*, proposes a prospective study on an alternative algorithm inspired from the analytical methods developed in Part III. In a nutshell, the algorithm simulates that users coordinatingly choose their departure time in order to get closer to an equilibrium situation. Its scope is restricted to networks with no tolls and it has been tested on a simple network. Although the method remains to be tested more extensively to control how it behaves on large networks, the first results are very encouraging.

Chapter 10, *An application to large interurban networks during summer holiday departures* presents a real-size application of the model on the French national road network (2404 arcs and 939 nodes). The model is used to assess an hypothetical time-varying pricing schemed intended to ease summer traffic congestion. The resulting computation time is perfectly acceptable and the qualitative analysis of the results shows the computed equilibrium gives consistent orders of magnitude. The numerical results indicate that even moderate variations of the toll with time might have strong impacts on an highly congested interurban network. By applying a time-dependant factor varying between 0.7 and 1.2 to the existing tolls, the aggregate travel times have decreased of approximately 10%.

# Acknowledgements

I would like to take this opportunity to thank my supervisor, Fabien Leurent, for all the many discussions and critique of my work throughout my time as his PhD student. His willingness to help, teach, and support me during this work, were invaluable. I am also highly indebted towards Vincent Aguiléra, who has been my secondary supervisor. I have Vincent to thank for his constant availability and his helpful advices in computer science. At the darkest times, Vincent always encouraged me: thank you also for that !

This work also owes much to the collaboration with Frederic Meunier, especially in its second part. Without Frederic's encouragements and his valuable mathematical skills, some chapters of this thesis would simply not have seen light.

I would like to thank Jerôme Renault and Guido Gentile for the attention the brought my work, and for their useful and kind commentaries. I would also like to thank Jean Delons and Jean-Patrick Lebacque, who kindly agreed to participate to my examination board, and Erik Verhoef, who accepted to preside over it.

I would also like to thank my friends and colleagues at the *Laboratoire Ville Mobilité Transport*. In particular, I would like to thank François Combes, Vincent Breteau, Caroline Guillot, Nicolas Coulombel, Olivier Bonin, Thierno Aw and Thai-Phu Nguyen.

My gratitude for the support, patience and trust of my family, notably during this work but not only, is beyond my capacity of expression. I trust they understand how much I owe them.

Finally my deepest thoughts go to Virginie, who supported me during these years of hard work. This thesis is dedicated to her.

# Introduction

## Dynamic user equilibrium in network assignment: operational and scientific context

**Operational stakes** Modern economy is based on trade between economic agents. People and firms trade for various goods, services or labour. This results in movements of goods and people. To be fulfilled, this demand for transport requires transport networks. Transport networks are characterized by a certain capacity, which corresponds to the number of passengers that can be transported per unit of time. When demand approaches capacity congestion may occur: travel speed, reliability and convenience decrease as the amount or length of trip-making increases. Although the proper evaluation of the costs of congestion is subject to vivid debates, there is a wide agreement that the economic stakes are considerable.

Increasing the capacity of the network to deal with congestion is one solution which needs to be assessed with a long term viewpoint. Investments in transport infrastructure and operating vehicles are long lasting expensive goods with sunk costs. Planning consists in designing, assessing and selecting the transport infrastructure.

A planned infrastructure is designed to cope with a certain volume of traffic. The actual volume of traffic might be larger as a result of unexpected growth in demand, or to day to day variations of the traffic volumes such as commuting peaks in urban contexts or seasonal peaks in interurban contexts. Transport authorities have at their disposal a wide range of short-term congestion management measures. These measures can be operational; some examples might include an increase in the frequency of service of public transport, the directing of traffic flows on alternatives routes or dynamic

speed control systems intended to homogenize traffic flow speeds. They can also be economic in nature, such as congestion tolls or modal shift incentives.

Transport capacity is a scarce resource in both time and space. Since massive investments in road infrastructure have decreased in recent years owing to their financial costs and environmental impacts, short-terms measures need to be optimized more than ever before. Fortunately there is space for improvements. Morning peaks can be spread out by using adequate time-varying pricing; the traffic on a congested route can be decreased by providing the right quantity and quality of dynamic information to users; adaptive traffic control systems can be installed at intersections to control changes in incoming flows traffic during the day. However, the challenges that need to be met are high. A deep understanding of the collective mechanisms leading to the allocation of capacity in time and space is necessary to correctly design these schemes. There is an urgent need for more detailed means to represent the interaction between travel choices, traffic flows, and time and cost measures in a temporally coherent manner.

**Scientific context** We have seen that congestion interferes with both short-term and long-term transport policies. Optimizing remedial measures requires a fine comprehension of the linkage between transport networks and transport demand. This is a difficult problem due to its numerous dimensions of complexity: traffic flows are the results of numerous trips with various origins and destinations; users' behaviours and preferences vary from one to another; congestion on the network results from spatially disaggregated interactions between users. For these reasons, finding a consistent solution – a (user) equilibrium in economic terms – requires considerable analytical sophistication.

The seminal work on the subject is by Beckmann, McGuire and Winston (1956) who showed how to find an equilibrium on an arbitrary transport network given certain assumptions. The operational tools they introduced, user equilibrium models (commonly known as *network assignment models*), are now widely spread in developed countries and used on a regular basis for traffic studies. Historically, the first assignment models had a static representation of transport in the sense that traffic flows and travel times were assumed to be constant over the simulation period.

This approach is problematic for two reasons. First, it ignores the dynamic aspects of congestion, *i.e.* the progressive accumulation and dissipa-

tion of large traffic volumes in certain areas of the network. This phenomena, sometimes referred to as *hypercongestion*, plays an important role in urban road transport. It is responsible for the essential part of the travel time losses. Second, it does not allow to model the time-varying aspects of either the origin-destination flows, dynamic traffic control measures, or time-varying pricing schemes.

The inherent limitations of the static assumption were soon identified. In this thesis, we will use the generic term of *dynamic network models* to designate all network models that represents variations in time of traffic flow and thus reflect the reality that transport networks are generally not in steady states. Here dynamic is a synonym for *time-varying*. Vickrey (1969) with his well-known bottleneck model, formulated a dynamic model of congestion on a single arc and establishes the resulting user equilibrium. Merchant and Nemhauser (1978) proposed a dynamic model of a transport network. The 1990's witnessed a renewed interest in dynamic user equilibrium models. The literature is divided into two different trends. On one hand, simulation-based models enhanced with equilibrium principles are developed (*e.g.* Dynasmart of Mahmassani, Hu and Jayakrishnan (1995)). On the other hand, the first rigorous analytical formulations of the dynamic user equilibrium appeared (Friesz, Bernstein, Smith, Tobin and Wie, 1993). At the end of that decade researchers began to realize that the equilibrium of the simulation-based models lacked some theoretical properties: their existence is not guaranteed, they are difficult to compute and are unstable. It is also around that time that the first large scale implementations of analytical dynamic user equilibrium models have appeared (Akamatsu, 2001; Bellei et al., 2005; Leurent, 2003*a*; Aguiléra and Leurent, 2009).

Analytical modelling of the dynamic user equilibrium is a rapidly evolving research topic. A large number of academic papers are focused on refining the physical representation of road traffic. The proper introduction of Daganzo's cell transmission model (Szeto, 2008) or the modelling of queue spill backs (Gentile, Meschini and Papola, 2007*a*) are examples of important theoretical and algorithmic advances. Yet the literature leaves a number of questions unanswered, especially regarding the dynamic representation of the transport demand. Some of them are stated hereafter.

- *The mathematical properties of user equilibriums remain to be fully established*: existence results for dynamic equilibrium models have been shown only in specific cases and no uniqueness and stability results

have been proven (Mounce, 2007). Along the same lines, mathematically concise formulations are rare, mainly due to the complexity of analytically formulating the traffic flow on a network.

- *User heterogeneity* is not fully represented: most of the existing analytical models consider a finite number of homogeneous groups of travellers, each group being characterized by a few variables *e.g.* vehicle type, value of time or preferred arrival time (for instance in De Palma and Marchal, 2002). A more general representation would be to consider continuous distributions over the space of characteristics. Although microeconomists have long considered small transport models with continuous heterogeneity (the bottleneck model is probably the most famous example), up to now no general theoretical formulation is available.

- *More efficient algorithms for user equilibriums with departure time choice* are still required. User equilibriums with route choice can now be computed with reasonable efficiency on real size networks (Aguiléra and Leurent, 2009). A substantial amount of work is still needed to properly state the appropriate numerical solution techniques when the problem includes departure time choice. From an algorithmic viewpoint this is probably one of the most challenging problems in transport science currently.

This list, although clearly not exhaustive, is sufficient to indicate that there is a need for a better representation of transport demand in dynamic user equilibrium models and more efficient algorithms to solve them.

## Problem statement

This thesis is focused on dynamic user equilibrium models for traffic assignment. It aims at providing formal properties and microeconomic foundations for computable DUE models.

This leads us to propose a mathematically rigorous and general formulation for the dynamic user equilibrium. Particular attention is paid to the representation of transport demand and more specifically to trip scheduling and users with highly heterogeneous preferences. This work is characterized by a high level of mathematical formalism; this allows for a precise, concise

and microeconomically consistent description of dynamic transport networks and dynamic transport demand. To the author point of view, the main contribution of the thesis is the formalization of dynamic user equilibrium models as Nash games.

Although the rigorous formalization of dynamic user equilibrium models is the main object of the thesis, it also deals with computational techniques of dynamic user equilibriums. We aim at solving analytically some stylized models to get a general intuition of the complex linkage between the demand and supply of transport in dynamic frameworks. The intuition acquired from solving these analytical models will be used to elaborate efficient numerical solving methods that can be applied to large size, real life networks.

Specifically, the work presented in this thesis has investigated these issues by studying the difficulties and the potential benefits of a finer representation of demand in dynamic equilibrium models. In particular, the thesis provides elements of answers to the following matters:

- *A theoretical framework for dynamic equilibrium models with continuous user heterogeneity.* Before aiming at fully operational and computable models, a formal framework for dynamic user-equilibrium models should be designed. Among the important theoretical questions is whether or not an equilibrium in such a model even exists.

- *What are the impacts of continuous user heterogeneity on dynamic user equilibriums?* How does this affect the physical distribution of traffic flows in time and space? What are the consequences in the cost distribution among users? To what extent are standard results altered by taking heterogeneity into account?

- *Evaluate the computability of dynamic user equilibrium models.* Obviously in the end our ability to correctly solve our model is essential. It requires the development of algorithms that (1) are effectively able to compute reasonable approximations of a dynamic equilibrium and (2) have reasonable requirements in term of computing times.

## Methodology

Our approach can be broken down into four steps. The first step is the literature review. Extensive reviews of academic works constitute the first stage of

this study. Second, a game theoretic formulation of the dynamic user equilibrium is proposed. It strongly relies on up-to-date results from mathematical economics on games with a continuum of players. Third, analytical resolutions of this model are presented in restricted cases. Although these specific cases are chosen to answer specific issues in transport economics, they gave us interesting insights regarding the mathematical structure of the problem. In particular they have been very valuable for the last step of this thesis, where a computable model is designed and corresponding solution methods are proposed.

This thesis heavily relies on mathematical techniques. In particular, great attention is given to the precise statements and mathematical correctness of the equilibrium we define. At first, some approaches and modelling choices might be perceived as unnecessarily sophisticated. Yet they have proven to be very useful and revealed that some commonly accepted assumptions are inconsistent.

Finally, our work was part of LADTA, a wider project led by the team Economie des Réseaux et Modélisation Offre-Demande of the LVMT [1]. LADTA, for **L**umped **A**nalytical **D**ynamic **T**raffic **A**ssignment, is a dynamic user equilibrium model introduced by Leurent (2003*b*) and designed as an extension of classical static assignment models, with special emphasis on the time-varying features. Recently the LTK (**L**adta **T**ool**K**it), a powerful implementation of LADTA main principles and associated solution methods has been developed by Aguiléra and Leurent (2009). One of the practical goals of this thesis was to enhance LADTA with a departure time model and to implement it in the LTK. For this last point we have benefited from the help of LTK's main contributor, Vincent Aguiléra. The relationship between LADTA and the thesis is two-sided. Obviously LADTA is a natural application of the general framework developed in the thesis. It has also been a useful case study upon which we have drawn in order to elaborate a more general theory.

## Outline

The thesis comprises ten chapters grouped into four parts.

- *Part I: Bibliography* is divided into two chapters. Chapter 1 deals with dynamic network modelling. It describes the physical mecha-

[1] Laboratoire Ville Mobilité Transport - UMR Ecole des Ponts ParisTech, INRETS, UPEMV

nisms leading to congestion. In addition, two important sub-problems are reviewed: dynamic least cost path and dynamic network loading. Chapter 2 covers the mathematical formulations and the computational methods of dynamic user equilibriums.

- *Part II: Dynamic congestion games and their application to dynamic traffic assignment.* In this part a new category of games intended to be a new framework for dynamic user equilibrium models is introduced. In Chapter 3 the game model is presented and two theoretical results are provided. First, a constructive proof of the existence and uniqueness of the solution to the dynamic traffic loading problem is exposed. Then a general existence theorem for Nash equilibriums in dynamic congestion games is given. In Chapter 4 it is established that the simplest dynamic user equilibrium model, known as the dynamic traffic assignment, can be seen as a particular case of dynamic congestion games. A known existence result, due to Mounce (2007), is then shown to derive from the existence result of Chapter 3. Most of the materials presented in Chapter 3 and 4 have been published in (Meunier and Wagner, 2010).

- *Part III: Analytical resolutions on simple cases.* Part III presents two simple dynamic congestion games for which the solutions can be derived analytically. The first game (Chapter 5) is a generalization of the Vickrey's bottleneck model as formalized in (Smith, 1984; Daganzo, 1985). Whereas Smith and Daganzo assume that the distribution of preferred arrival time is S-shaped, we consider more general distributions. This leads to a much more complex pattern of congestion and in particular gives insights on the way companies' work schedules can impact morning peak hours. The results of Chapter 5 have been published in (Leurent and Wagner, 2009). The second game (Chapter 6) models a two-route tolled network where users are continuously heterogeneous with respect to their value of time. This allows us to conduct a study on the relative efficiencies of various pricing strategy and how it is affected by the level of heterogeneity in users' value of time.

- *Part IV: Numerical methods for the dynamic user equilibrium with departure time choice.* In Chapter 7 a simple dynamic user equilibrium model is stated in the formalism of dynamic congestion games. Chapter 8 is devoted to the presentation and assessment of an algorithm

inspired from the classical convex combination scheme. Chapter 9 proposes a prospective study on an alternative algorithm inspired from the analytical methods developed in Part III. Finally, Chapter 10 presents a real-size application of the model on the French national road network. Part of the results presented in Chapter 10 were published in (Aguiléra and Wagner, 2009).

# Part I

# Bibliography

# Part introduction

## What is traffic assignment at equilibrium?

In the transport field, the most basic user equilibrium problem can be informally stated as follows.

*Given:*

- *A transport infrastructure supply represented by a network consisting of nodes and arcs. With each arc is associated a way to represent congestion (a congestion model), i.e. some function or procedure that allows deriving the travel time on an arc knowing the flow of travellers that goes through it.*
- *The travel demand modelled by an origin-destination (OD) matrix with network users' departure rates from each origin node to each destination.*

*find the users' flows and the travel times on each the network arcs.*

This problem is known as that of *traffic assignment*, since the issue is how to is to assign the OD matrix onto the network. To solve the traffic assignment problem, it is required that the rule by which network users choose a route be specified. This rule can be viewed as the function or the procedure that specifies the demand for transport over routes. The interaction between the routes chosen between all OD pairs, on the one hand, and the congestion

models on all the network arcs, on the other, determines the equilibrium flows and corresponding travel times throughout the network.

The transport infrastructure is typically a network of motorway segments and the network users' road vehicles. But one might consider traffic assignment for all sort of transport infrastructures, such as transit network, and for all sort of network users, such as pedestrian or cyclists.

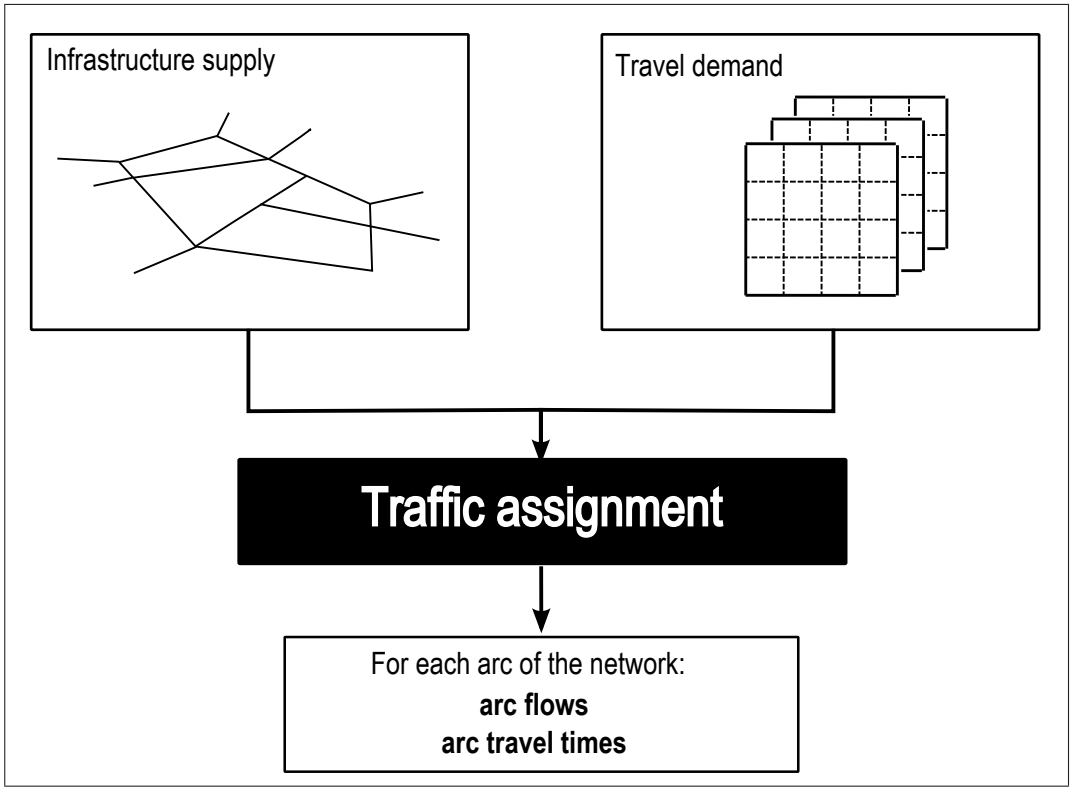



Figure 1: General framework of a traffic assignment procedure

This rule is usually derived by assuming that every network user will try to minimize his own travel time when travelling from origin to destination. This does not mean that all users between each origin and destination pair should be assigned to a single route. The travel time on each arc changes with the flow and therefore, the travel time on several routes changes as the arc flows change. A stable condition is reached only when no user can improve his travel time by by unilaterally changing routes. This is the characterization of the user-equilibrium (UE) condition, well known as the Wardop's principle.

## Static and dynamic traffic assignment

There is a fundamental distinction between the congestion models used for traffic assignment: they can be either static or dynamic.

- In static congestion models, travel time and users' flows are assumed to be time independent. Thus a static congestion model is usually a simple function that maps a user flow with a travel time. This function is often referred to as an arc performance functions.

- In dynamic congestion models, users flows and travel times may vary with time. Consequently the range of congestion models is much wider.

Although less realistic, static congestion models are especially convenient because they lead to well-posed traffic assignment problems for which they exist powerful computational techniques. They are now used on an everyday basis by most transport policy analysts.

Until recently dynamic models were essentially research objects. Indeed the resulting assignment problems, *dynamic traffic assignment problems* are relatively more complex to formulate and to compute.

## Dynamic user equilibrium problems

The problem of dynamic traffic assignment as presented above implicitly assumed that users may only choose their route. Now, with dynamic congestion models, travel times vary so network users (acting as rational agents) should also be able to choose their departure time. Starting from this observation, more general dynamic user equilibrium (DUE) models have been elaborated.

Such models are known as *DUE models with departure time choice*. In the resulting equilibrium problems, there are no longer OD matrices as users' departure rates now change according to users' departure time decisions. That's why we will no longer speak of dynamic traffic assignment but rather of DUE problems[2].

[2]Note that some authors speak of dynamic traffic assignment with departure time choice. In this thesis, we have chosen not to use this terminology.

## Scope of the bibliographic review

This thesis studies DUE models with departure time choice. Our first step is to provide a mathematically rigorous, microeconomically consistent and general formulation for these models. The bibliographic review will provide materials for this task. Consequently, we will focus on the following aspects:

- **Dynamic models**. We do not review static models of the user equilibrium. Indeed the subject is extremely wide, already well covered and we have felt that it would not bring much to our matter.

- **A microeconomic and mathematical viewpoint**. The approach chosen here is not the one of traffic engineering, even if a significant amount of the literature reviewed comes from this field. Here, the attention is rather paid on the way users decision process are represented and formalised.

- **Analytical congestion models**. The traffic phenomenons that lead to congestion can be modelled using analytical, simulation-based or even statistical methods. We essentially cover analytical approaches. Moreover, the focus is not on the way the traffic is represented but we rather look at the general properties of the congestion models and their consequences on the equilibrium structure.

- **Deterministic models**. Although there is an increasing number of works on stochastic DUE models, we do not review them. Indeed, the rest of thesis does not deal with stochastic issues in DUE models[3].

## Structure of the part

This part is divided in two chapters. Chapter 1 is devoted to the supply side of a DUE model. It presents models of congestion on a network. Although it focuses mainly on road congestion, other from of congestion are reviewed (notably parking and public transport). This chapter also deals with two types of algorithms on a dynamic transport network, namely least cost route

[3]However a careful reader will realize that, when considering continuously heterogeneous demand, the differences between deterministic and stochastic DUE models are very thin.

algorithms and network loading algorithms. To present them in a unified fashion, we introduce a common framework for dynamic transport networks.

Chapter 2 then presents equilibrium models that can apply to dynamic transport networks as stated in Chapter 1. This chapter first presents different formulations of the dynamic traffic assignment problem. All of them are expressed under a unified notation framework and the equivalence results between each formulations are given. The chapter then summarizes different works on more general DUE models.

## Some vocabulary

Terminology regarding DUE models may vary from one author to another. Thus we have felt it was necessary to provide the following definitions. Although some of them are shared by several authors, they have no universal meaning outside the scope of this thesis.

**Network users** Even though most of this thesis refers to road traffic, the generic term of "(network) user" is retained. A user might represent all kinds of entities (pedestrians, cyclists, travellers, vehicles, trucks, ...).

**Congestion model** Informally, an arc congestion model is a mathematical object that encompass the traffic phenomena that lead to congestion. Precisely, in this thesis, an arc congestion model is a mapping between a time-varying flow of users and a time-varying travel time. A similar definition could be introduced for intersection (or node) congestion.

**Dynamic transport network** Informally a dynamic transport network is a mathematical object that models a transport infrastructure and the congestion phenomenons that might occur on it. Precisely, in this thesis, a dynamic transport network is modelled by a graph where each arc (and possibly each node) is associated with an arc (or a node) congestion model.

**Dynamic Wardrop assignment** A DUE problem with only route choice is refer to as a dynamic traffic assignment problem. The dynamic traffic assignment according to the Wardrop principle is referred to as the dynamic Wardrop assignment.

Chapter 1

# Dynamic network modelling and algorithmics

A simple definition of the supply of transport between an origin and a destination is the set of available *transport services* serving this origin-destination pair, a transport service being defined as a departure time, a mode and a route. The assessment of a transport service by a user depends on the expected travel time proposed by a transport service and on its monetary costs. Now there are other characteristics that might be taken in account, typically the reliability of the travel time, convenience of travel, and the expected arrival time. The set of such characteristics represents the *quality of a transport service*.

Transport occurs on networks and thus transport supply is intricately related to transport network modelling. A service's travel time depends on several factors including the transport network structure and its operating rules as well as the traffic load. The influence of the latter on travel time is called congestion. The objective in this chapter is to analyse the determinants of travel time, travel costs and congestion in dynamic (*i.e.* time-varying) frameworks, assuming a given network structure.

This chapter treats of transport network modelling and of the associated algorithmic issues. It is organized as follows. Section 1 gives the basic principles of congestion and cost modelling in dynamic transport networks at the elementary level of an arc then a junction. Section 2 defines formally the concept of dynamic transport network and gives some notations. Two subsequent sections deal with two problems that pertain to dynamic transport networks. The first one, known as the *continuous dynamic network loading problem*, consists in determining the traffic flow propagation in a network

subject to congestion. The second problem is that of *shortest route and least cost route problems* in time-varying networks.

# 1 Congestion dynamics and user costs at the elementary level

A general economic definition of a congestion prone facility is that the quality of service decreases with the intensity of use. For transport, road congestion is perhaps the best illustration. Yet congestion in transport is not limited to road: it also affects travellers on bus and subway networks. Even on pavements some forms of pedestrian congestion might occur. Now we will only briefly mention public transport and essentially focus on road congestion.

Road congestion arises from many physical mechanisms. A classic distinction is between flow and bottleneck congestion. *Flow congestion* arises from the local interactions between drivers: slower cars are getting in the way of faster cars, drivers can't adjust their speed instantaneously... *Bottleneck congestion* arises when a drop in the capacity somewhere on the road network causes traffic queues to form. A related distinction is between arc and intersection (or nodal) congestion. This latter is quite appealing when representing transport supply as a transport network and we will retain it in our exposition.

In the first two subsections we thus present arc and then intersection congestion. These two subsections solely focus on the effect of congestion travel time. In the subsequent subsections are listed some other relevant topics about congestion modelling.

## 1.1 A preliminary remark on the mathematical nature of traffic flows

In a dynamic transport network model, the basic quantities are *time-varying flows of users*. Although most of this thesis refers to road traffic, the generic term of "user" retained here might represent all kinds of entities (pedestrians, cyclists, travellers, vehicles, trucks, ...). A time-varying flow can be represented as a map from a set of clock times (instants) $h \in \mathcal{H}$ to the set of positive real numbers. Denote it $h \mapsto x(h)$. Now not all such maps are representing physically sound flows. A natural requirement on users' flows

is to be integrable on every bounded susbsets of $\mathcal{H}$. Then integrating the map $x$ over an interval $I$ of $\mathcal{H}$ gives the number of users that went through a certain point in space during $I$. When $x$ is not integrable this quantity is not necessarily defined which is difficult to interpret physically. Consequently the integrability of $x$ is a physical requirement.

$x$ being integrable, one defines its corresponding *cumulated flow* $X : h \mapsto \int_{h_m}^{h} x(u)\ du$ for all $h > h_m$, where $h_m$ is some reference instant. The interpretation of $X(h)$ is straightforward: it is simply counting the number of users that went through a given point between $h_m$ and $h$. Using cumulated flows, instead of "normal" flows (call them *instantaneous flows*), is convenient to express conservation laws. A natural question arises: what is the set of cumulated flows that corresponds to physically sounded instantaneous flows? The answer to that question is given by standard real analysis results. Assume $\mathcal{H}$ is a bounded interval $[h_m, h_M]$ and a function $X$ on $[h_m, h_M]$ such that $X(h_m) = 0$; then there exists a (Lebesgue) integrable function $x$ such that $X(h) = \int_{h_m}^{h} x(u)\ du$ if and only if $X$ is absolutely continuous [1] [2] (see for example Rudin, 2009).

For a given cumulated flow $X$ there might be several possible instantaneous flows $x$ but they are equal almost everywhere. Thus we will consider the set of integrable functions from $\mathcal{H}$ to $\mathbb{R}_+$ quotiented out by the equivalence relationship "equal almost everywhere"[3]. It is denoted $L(\mathcal{H}, \mathbb{R}_+)$. For any increasing and absolutely continuous function $X$ from $\mathcal{H}$ to $\mathbb{R}_+$, there is a unique corresponding instantaneous flow $x \in L(\mathcal{H}, \mathbb{R}_+)$. It is possible to create a bijection between instantaneous flows (taken from $L(\mathcal{H}, \mathbb{R}_+)$) and cumulated flows (taken from the set of increasing and absolutely continuous functions). For this reason the set of increasing and absolutely continuous functions on $\mathcal{H}$ is abusively denoted $L(\mathcal{H}, \mathbb{R}_+)$. To avoid confusion the following convention is adopted: a cumulated flow is always denoted by a capital letter while the associated instantaneous flow is denoted by the corresponding

[1] A function $F : [a, b] \to \mathbb{R}$ is absolutely continuous if for every *epsilon* $> 0$, there exists $\delta > 0$ such that for all sequences $([a_n, b_n])_n$ of disjoint intervals of $[a, b]$: $\sum_{n\geq 0} (b_n - a_n) < \delta \Rightarrow \sum_{n\geq 0} |F(a_n) - F(b_n)| < \epsilon$. This definition is equivalent to $F$ has a derivative $f$ almost everywhere, $f$ is Lebesgue integrable, and $F(x) = f(a) + \int_a^x f(t)\ dt$ for every $x \in [a, b]$.

[2] Note that the set of functions differentiable almost everywhere can not be used here as a function might be differentiable almost everywhere with a derivative that is not integrable.

[3] In other terms we consider the set obtained by identifying the elements $f$ and $g$ such that $f$ equals to $g$ almost everywhere.

lower-case letter. For instance if $X$ is a cumulated flow then $x$ is the corresponding instantaneous flow and consequently $x = \frac{dX}{dh}$ almost everywhere. Then $x \in L(\mathcal{H}, \mathbb{R}_+)$ is read "$x$ is an integrable functions $\mathcal{H} \mapsto \mathbb{R}_+$" while $X \in L(\mathcal{H}, \mathbb{R}_+)$ is "$X$ is an increasing and absolutely continuous function on $\mathcal{H}$".

## 1.2 Road congestion on arcs

To stay as general as possible, let us define an arc congestion model as a mapping between a time-varying flow and a time-varying travel time. There are two common requirements on a congestion model. The first one, namely the *causality principle*, states that the arc travel time of a user entering at an instant $h$ solely depends on the flows entered before $h$. The second one, the *FIFO principle*, states that users exit the arc in the same order they entered it. The FIFO principle is sometimes translated as a forbidden overpassing rule. For flows of homogeneous users it states in fact much more: that it is inconsistent that the *same* user might arrive earlier at arc's exit by entering later on this arc. These two principles are of great help in the assessment of the models presented below.

In this subsection, we overview different approaches for the modelling of time-varying congestion on a single arc. We first limit ourselves to flows of homogeneous users and then briefly consider multi-class users flows.

**Hydrodynamic models.** Hydrodynamic models are traffic models inspired by fluid mechanics. The most widely used hydrodynamic model was developed by Lighthill and Whitman (1955) and Richards (1956) and is known as the LWR model. In the LWR model the traffic streams on an arc are represented by time- and space-varying traffic flows, densities and speeds. It assumes that the speed-density relationship embedded in the fundamental traffic diagram also holds under non-stationary conditions at every point of space and time. The model is closed by a traffic conservation law. It leads to the formulation of a single partial derivative equation (see Frame 1 for details).

We denote $\rho(h,s)$, $x(h,s)$ and $v(h,s)$ the density, flow and velocity at time $h$ at a point with the curvilinear abscissa $s$, and $f$ the speed-density relationship. First, for physical consistency, the following relationship is required:

$$x(h,s) = v(h,s).\rho(h,s)$$

Then the two assumptions of the LWR model write down as:

$$\frac{\partial x(h,s)}{\partial s} + \frac{\partial \rho(h,s)}{\partial h} = 0 \quad \text{(conservation law)}$$

$$\rho(h,s) = f(x(h,s)) \quad \text{(speed-density relationship)}$$

This yields the following (well-known) partial derivative equation:

$$\frac{\partial f\big(\rho(s,h)\big)}{\partial s} + \frac{\partial \rho(h,s)}{\partial h} = 0 \tag{1.1}$$

The LWR model is especially useful for the study of traffic shock waves. Figure 1.1 gives a simple example. It is a time-space diagram showing the trajectories of representative vehicles. Initially, the arc is in stationary state described by a density $\rho_d$ and a flow $x_d$. If the inflow at entrance changes to $x_u$, this results, through the flow-density relationship, in a new density $\rho_u$ at entrance that will propagate downstream as a shock wave. According to the conservation law implies that the speed is determined by the Rhankine-Hughoniot formula:

$$v_{wave} = \frac{x_d - x_u}{\rho_d - \rho_u}$$

Frame 1: The LWR model

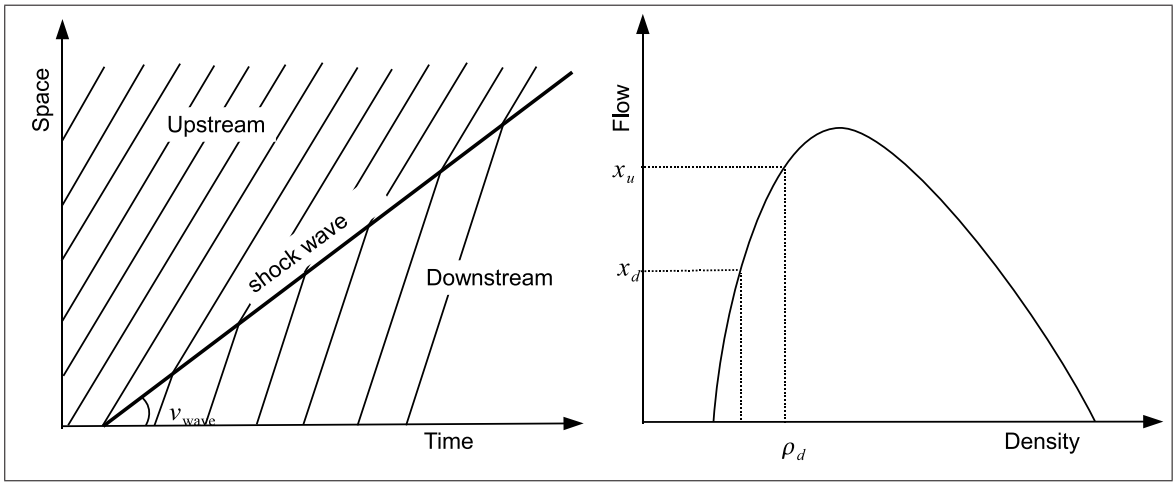



Figure 1.1: Shock waves with the LWR model: flow density relationship (right) and time space diagram (left)

The model is analytically untractable in the general case, but the partial derivative equation can be solved numerically by classical partial derivative equation techniques such as the Godunov's scheme (Lebacque, 1996). Now the most widely used solution is the cell transmission model proposed by (Daganzo, 1994) as an independent approach, that can be shown to be a spatial and temporal discretization for the LWR for the following flow-density function:

$$f(\rho) = \max\left\{v.\rho, x_{\max}, v.(\rho_j - \rho)\right\}$$

where $v$, $x_{\max}$ and $\rho_j$ are parameters that can be interpreted respectively as the free flow speed, the maximal flow and the jam density. The success of the cell transmission model is probably due to the simplicity of its implementation and its reasonably good computational properties. However the model is still tedious and computationally demanding on large networks with arcs with large spatial dimensions. It requires a dense spatial discretization, in the sense that each road link is represented as a important sequence of cells (a typical space discretization step is 50 meters). Therefore various simplifications have been developed, some of which are presented in the next paragraphs.

Although the LWR model can predict some traffic phenomena rather well, it is also known to have some flaws. Its main restrictive assumption is that the speed-density relationship holds exactly at each point in time and space regardless of the possible drivers anticipations. This leads to instantaneous

speed adjustment and thus to infinite acceleration. Among the known consequences is the impossibility to represent stop and go waves. In an attempt to correct those flaws higher order differential equations have been proposed, starting in the early 70s (Payne, 1971).

**Bottleneck models.** The *pointwise bottleneck model* was introduced by May and Keller (1967) and is famous for its use by Vickrey (1969).

The bottleneck model assumes that travel on the arc is uncongested except perhaps on a single bottleneck of deterministic capacity $k$. If the incoming flow at the bottleneck exceeds $k$ a queue began to form and users have to wait according to a FIFO discipline before leaving the bottleneck. The analytic of the model, as presented in (Arnott, De Palma and Lindsey, 1998), writes down as follows. Assume an inflow $x$ and denote $h \mapsto t(h)$ the resulting travel time. Then the function $t$ writes down as below, where $Q(h)$ is the number of users queuing at $h$ in the bottleneck, and $t_0$ is the free flow travel time:

$$t(h) = t_0 + \frac{Q(h)}{k} \tag{1.2}$$

where $Q$ stems from the following differential equation:

$$\frac{dQ}{dh}(h) = \begin{cases} x(h) - k + \dfrac{Q(h)}{k} & \text{if } Q(h) \neq 0 \text{ or } x(h) - k > 0 \\ 0 & \text{otherwise} \end{cases} \tag{1.3}$$

Here the capacity is to be understood as the maximal flow that can go through the arc. The restriction in capacity can arise from many causes: the geometry of the roadway, speed restriction, lane reduction or an intersection limiting the capacity at the extremity of the arc. Note that the latter case is related with the next subsection dealing with congestion at intersections.

In our physical description of the model, we stated the queue was punctual and this features gives its name to the model. More realistic models of queuing have been developed, often termed as *physical queue models* or *horizontal queue models*. They are compatible with the LWR model with the assumption that the flow-density relationship has a triangular shape in the bottleneck area and allow to estimate the physical extent of the traffic queue. This latter feature is particularly important when one wants to model queue spill back on neighbouring arcs.

Bottleneck models respect by construction both the causality and the FIFO principles.

**Flow-delay and volume-delay models.** The two previous approaches described traffic congestion by an explicit model of the traffic flowing over the arc. Another widely used approach is to consider that the travel time on an arc when entering at an instant $h$ is some function of the characteristics of the arc at $h$ (*e.g.* Ran and Boyce, 1996). The rationale behind these models appears to be an attempt to generalize the classic performance models used in static transport network models. Two variants exists.

- In *flow-delay* models, the travel time is taken as a function of the instantaneous flows at the time of entrance.

- In *volume-delay* models, the travel time is taken as a function of the traffic volume on the arc at the time of entrance. The volume-delay function can typically be derived from standards speed-density relationships by considering the average density on the arc. To the author's knowledge they were introduced by Janson (1991).

Although the decision of which function to retain would typically imply some field measurements, the great majority of the literature simply assume some form of BPR type relationships without further justifications. By definition flow-delay and volume-delay models respect the causality principle. However, it is easy to see that they do not always satisfy the FIFO principle. For volume-delay models, there are two trends in dealing with this problem. First, one can assume that volume-delay functions are linear (Daganzo, 1995). Then, any arc cumulated volume leads to a FIFO travel time function. Second, one can assume some more advanced conditions on the maximum variation in route cumulated flows. In both cases these methods strongly limit the scope of the model.

A physical interpretation of flow-delay model is that shock wave travel at the same speed as vehicles and therefore never influence other vehicles. This is why they are sometimes refer to as *no-propagation models*.

**Exit flow models** Introduced by the seminal paper of Merchant and Nemhauser (1978), exit flow models assume that the outflow of an arc solely depends on the traffic volume on this arc or equivalently on the average density. Assuming that vehicles travel in a FIFO manner, the corresponding travel times can be easily derived.

A physical interpretation is that an exit flow model entails this assumption that density remains uniformly distributed over the arc. Thus an increase in inflow immediately results in a corresponding increase in the density along the arc. It implies that shock waves propagate at an infinite speed. This is why some refers to exit flow models as *instantaneous propagation models* (Lindsey and Verhoef, 1999). Note that in an exit flow model vehicles might be affected by traffic behind them and thus exit flow models violate the causality principle.

**Comments on microsimulation models.** The primary focus of this thesis is about analytical models of the dynamic user equilibrium, so microsimulation models are only briefly reviewed. Microsimulation models are sometimes called microscopic models or vehicle-based as they explicitly describes the motion of each vehicle as opposed to the former methods which are macroscopic or flow-based models.

The basis of nearly every microscopic models is a car-following model, as developed from the late 50s up to the early 60s upon an original prototype by Chandler, Herman and Montroll (1958). In a car-following model the motion of a vehicle is a function of the motion of the vehicle immediately ahead. A simple formulation is given by the following differential equation:

$$a(h) = c\frac{\Delta v(h-T)}{\Delta x(h-T)} \tag{1.4}$$

where $a(h)$ is the of the following vehicle at instant $h$, $T$ is a reaction time, $\Delta v$ is the difference of speed between the two vehicles, $\Delta x$ is the spacing between them and $c$ is a non-negative parameter. Equation (1.4) states that the acceleration of a vehicle is proportional to what can be interpreted as the temporal distance from the vehicle ahead. An interesting feature of car-following models is that they imply, under suitable assumptions, the LWR model at a macroscopic level. Now they can easily be extended to better fit with real models by recognizing that vehicles accelerate at finite rates or that they anticipate future traffic conditions, whereas in the LWR the right approach to do that is still under discussion.

Modern microsimulation models are considerably more complex than a simple car-following model. They commonly integrate some stochastic components, integrate overpassing models or elaborate engine models (Algers, Bernauer, Boero, Breheret, Di Taranto, Dougherty, Fox and Gabard, 1997).

The possibilities are virtually infinite but there is a real need for correctly assessing these models with respect to field measurements that has not been entirely fulfilled yet.

**Comments on multi-class congestion models.** As mentioned earlier, the models presented above are valid for flows of homogeneous users. By homogeneous users it is meant users with the same driving behaviour and whose vehicle have similar physical characteristics.

Now vehicle types can be distinguished according to their difference in size, maximum speed, acceleration and deceleration rate. Almost all microscopic simulation models distinguish several vehicle types for instance trucks, passenger cars or motorcycle (Algers et al., 1997). In that order of idea an hydrodynamic model considering passenger cars and trucks is presented in (Hoogendoorn and Bovy, 2000).

Another acknowledged fact is that not every person acts in the same way in terms of traffic behaviour. Some drivers might be aggressive and wish to drive at higher speed than more cautious ones and this has consequences on the way traffic flow. In the simulation model PARAMICS (Smith, Duncan and Druitt, 1995) the characteristics of different drivers within the network are determined by allocating random values of aggression and awareness to the driver of each vehicle. Many other simulation models exist, all making their own distinctions in driver characteristics and/or vehicle type characteristics. When it comes to analytical approaches, Ran and Boyce (1996) describe a macroscopic model where network users are stratified based on driver characteristics, such as driving behaviour (cautious, rushed, ruthless), on driver's income and age, or on route diversion willingness (one route, few alternative routes, en route diversion).

## 1.3 Road congestion at intersections

Congestion at intersections is a complex matter that has received a lot of attention from traffic engineering and an exhaustive review is beyond the scope of this thesis. Here we simply review some simple determinist approaches, and summarize the physical mechanisms undermining intersection's congestion.

**Unsignalized intersections.** When a set of flows meet at an intersections there is a competition for the limited amount of capacity available. A popular model of unsignalized intersections has been developed by Daganzo (1995) as part of its cell transmission model. The model treats two specific cases of intersections, with either two incoming arcs and one outgoing (*a converge*) or two outgoing arcs and one incoming (*a diverge*). In Frame 2, Daganzo's main assumptions of for both types of intersections are presented.

In a nutshell, the converge model is ruled by priority coefficients that specify the minimal share of the junction capacity that is aloted to each flows while the diverged model relies on a FIFO principle and capacity limits at the entry of the outgoing arcs. Daganzo's model was extended in (Durlin and Henn, 2008) to deal with general intersections with any numbers of incoming and outgoing arcs.

**Daganzo's intersection model**

- Consider the converge described in Figure 1.2. The flows arriving at the two incoming arcs' tails are $x_1^+$ and $x_2^+$ and we wishes to determine what are the flows $x_1^-$ and $x_2^-$ that will actually enter in the intersection. The differences between the flows arriving and the ones actually entering remains on the corresponding incoming arcs. A convergent intersection is characterized by a capacity $k_{int}$, that represents the maximal flow that can go through the intersection. If $x_1^+ + x_2^+ > k_{int}$ then it is necessary to determine the proportions $k_{\text{j}}$ that will be assigned to each arc. This is achieved by introducing two reals $\alpha_1$ and $\alpha_2$ such that $\alpha_1 + \alpha_2$ representing the priority of each flows with respect to the other one. More precisely $\alpha_i$ represents the alloted share of the capacity to flow from arc $i$. This yields the following rule: $x_1^+ > \alpha_i k_{int}$ implies $x_1^- \geq \alpha_i k_{int}$.

- Consider the diverge described in Figure 1.2. The incoming flows are described by $x_1^+$, the flow of users wishing to enter arc 1 and $x_2^+$ describing the one wishing to enter arc 2. Our aim is to determine what are the flows $x_1^-$ and $x_2^-$ that will actually enter in the junction. The difference between remains on the incoming arc. A convergent junction is characterized by the capacities $k_1$ and $k_2$ of the entrance of arc 1 and 2, that represents the maximal flow that can go inside each arcs. Daganzo's main assumption is that vehicles are unable to exit prevent all those behind, regardless of destination, do continue. That is to say that users wait under a FIFO discipline at the diverge. Mathematically that is to say that $x_1^+/x_2^+ = x_1^-/x_2^-$. Incorporating the two physical constraints yields the compact formulation: $x_1^- = \min\left\{x_2^+, k_1, \frac{x_1^-}{x_2^-}.k_2\right\}$ and $x_2^- = \min\left\{x_1^+, k_2, \frac{x_2^-}{x_1^-}.k_1\right\}$.

Frame 2: Daganzo's intersection model

To our knowledge no precise analysis of the *intersection's congestion externalities* have ever been conducted. This is quite surprising as intersection congestion is known to be predominant over arc congestion in urban contexts. Now Daganzo's model gives interesting economic insights regarding this topic. Consider a set of flows of vehicles arriving on a congested intersection from arcs $i \in I$ and leaving the intersection from arcs $j \in J$. If one of them, call it $x_{ij}^+$ slightly increases, then the waiting time to enter the intersection may increase in return for all of them through two effects. If $x_{ij}^+$ is smaller than $\alpha_i k_{int}$, than the amount of capacity aloted to $x_{ij}^+$ might increase, thus reducing the capacity available for the other flows. This is the *converge's part* of the intersection's congestion externalities. Then if $x_{ij}^+$ is bigger than $k_j$, this will slow down the overall flow on the intersection through the FIFO waiting discipline. This is the *diverge's part* of the intersection's congestion externalities.

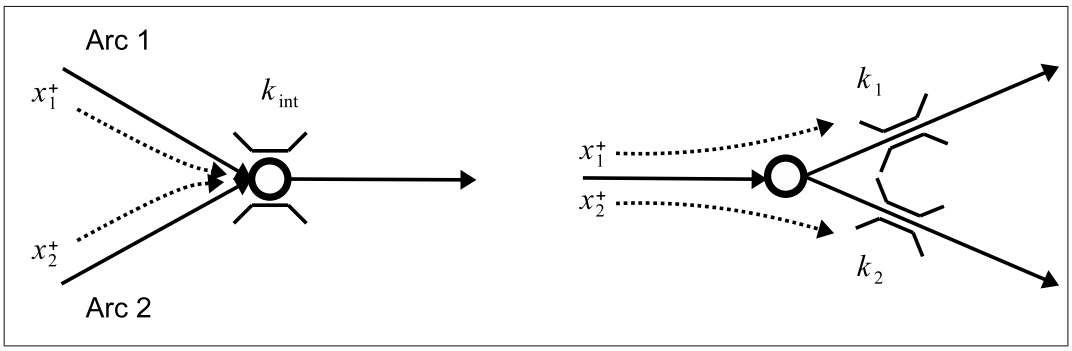



Figure 1.2: Converge (left) and diverge (right) in Daganzo's model

**Signalized intersections.** In signalized intersections, the capacity assigned to each flows is mostly predetermined by the traffic signal plan. Thus the intersection simply imposes a capacity on the arc's exit and from a modelling perspective it can be represented directly in the arc congestion model. The capacity assigned to each arc might be consider as constant or time-varying in order to represent finely the succession of green and red phases.

**Spillback congestion.** When the traffic repartition on the arc is explicitly represented, for instance by LWR or horizontal bottleneck models, there might be situations when the incoming traffic on an arc is limited because

of a hight traffic density at the head of it. Such a phenomenon is known as the queue spillback. Several models incorporate such a feature, which is known to be qualitatively important (Adamo, Astarita, Florian, Mahut and Wu, 1999; Gentile, Meschini and Papola, 2007*b*). The traditional approach is simply to use the unsignalized intersection model and to set the arc entry capacities accordingly.

## 1.4 Other forms of congestion

**Congestion in public transport.** Congestion in public transport is present under various forms. Leurent (2010) reviews some of them by identifying the various capacity bottlenecks occurring in the public transport system. Among others, Leurent identifies *vehicle related capacities* (maximum number of passengers per vehicle, number of seats, maximum number of passengers that can board during a stop), *station related capacities* (maximum of passengers per platform, maximum of flows of passengers in the corridors), and *mission related capacities* (maximum number of passengers per mission). When one of those capacities is overflowed by the number of passengers, it results in some form of congestion, either by an increase in the waiting time, in passenger's discomfort (*e.g.* seat congestion) or in service disturbances resulting in a global increase in the travel times.

The representation of public transport in dynamic transport network models is a recent and non mature topic. The least cost route algorithm introduced by Ziliaskopoulos and Wardell (2000) provides an interesting framework to represent route that accounts for both highway and transit mode. When it comes to dynamic user equilibrium models, the only form of congestion modelled is the one arising from the limited capacity of vehicles, that might cause queuing. Examples of such models are presented in (Tong and Wong, 1999; Tong, Wong, Poon and Tan, 2001; Nguyen, Pallottino and Malucelli, 2001).

**Parking congestion and costs.** Parking is important for a number of reasons. The monetary costs of parking, when it is not freely provided or strongly subsidized, represents a large part of the total monetary costs of a car trip. Finding cheap and convenient parking spaces typically entails

cruising for parking and this contributes to traffic congestion[4]; when the free spaces are rare, this search for parking spaces might represents a significant part of the total travel time. Street parking also interacts with traffic flows in a complex ways resulting in capacity drops.

Recently some theoretical works focused on the externalities generated by parking decisions (*e.g.* Anderson and de Palma, 2004). A parking externality arises because individuals neglect the increase their parking causes on the mean density of occupied parking spaces and thus on the average parking searching time.

Parking congestion essentially results from the accumulation of parked vehicles. It is essentially a stock congestion and consequently the correct framework to represent is dynamic. To our knowledge, only a few studies deals with parking, and all of them are simulation-based.

# 2 A model of dynamic transport network

Now that the main modelling approaches in transport networks have been reviewed, let us introduce a few notations. They will be of great use for the rest of the chapter.

**Time** Time variables will be denoted in two different ways. $h$ represents a clock time and $t$ a travel time (or more generally a difference between two clock times). The allowable departure time period for the users is a bounded interval $\mathcal{H} = [h_m, h_M]$ although the time period under study will be a longer time interval $\bar{\mathcal{H}} = [h_m, \bar{h}_M]$. Note that $\mathcal{H}$ and $\bar{\mathcal{H}}$ have the same initial instant.

**Network topology** Let $(N, A)$ denote a transport network composed of a set $N$ of nodes $n$ and a set $A$ of arcs $a$. Let $OD \subseteq N \times N$ denote the set of origin-destination (O-D) pairs. A route $r = a_1, \ldots, a_n$ is a sequence of arcs without repetition. The notation $r \prec a$ (resp. $r \preceq a$) represents the sub-route composed of the arcs in $r$ before $a$, $a$ excluded (resp. $a$ included). $R_{od}$ is the set of routes connecting the origin of the

[4] Shoup (1997, Table 11-5) displays the results of 16 studies on cruising for parking in downtown cities. The mean share of parking cruising among the total traffic flow was 30% and the average search time was 8.1 minutes. While the study locations were not chosen randomly, the results still indicate the importance of cruising for parking.

O-D pair *od* to its destination. The set $R := \cup_{od \in OD} R_{od}$ is the set of routes.

**Route and arc flow vectors** Two types of (cumulated) flows will be considered, *route cumulated flows* and *arc cumulated flows*. Route cumulated flows are defined on $\mathcal{H}$ while arc cumulated flows are defined on $\bar{\mathcal{H}}$. A vector $\boldsymbol{X}_R = (X_r)_{r \in R}$ is a *route flow vector* while a vector $\boldsymbol{Y}_A = (Y_a)_{a \in A}$ is an *arc flow vector*. Recall that the formal definition of cumulated flows is exposed in 1 and that their set is denoted $L(\mathcal{H}, \mathbb{R}_+)$. A useful operation on flows is their *restriction on* $[h_m, h]$: for any $h \in \mathcal{H}$ and any flow $X$, the quantity $X|_h$ is a cumulated flow defined on the same interval than $X$ and such that $X|_h(u) = X(u)$ for $u < h$ and $X|_h(u) = X(h)$ otherwise.

**Travel time and exit time functions** A travel time function is a function of the (clock) time that gives the travel time on an arc when entering a route or an arc at a given instant. An exit time function gives the exit time for a given entrance time. Travel time functions are denoted $h \mapsto \tau_a(h)$ (travel time on arc $a$) and $h \mapsto \tau_r(h)$ (travel time on route $r$) while exit time functions are denoted $h \mapsto H_a(h)$ and $h \mapsto H_r(h)$.

**Arc travel time model** An arc travel time model is a function from the space of arc cumulated flows to the space of travel time functions. It maps a time-varying flows defined for every $h \in \bar{\mathcal{H}}$ to a travel time function defined on $h \in \bar{\mathcal{H}}$. In other words, an arc travel model is simply a compact notation to describe the physical phenomena underlying traffic congestion. It can represent most of the models presented in the previous section. Arc travel time models are denoted $Y_a \mapsto t_a[Y_a] = \tau_a$.

With these notations, it is easy to formalize the causality and FIFO principles that were qualitatively exposed earlier.

**Definition 1.1** (Causality principle)**.** *An arc travel time model is said to obey the causality principle if:*

$$t_a[Y_a](h) = t_a[Y_a|_h](h) \quad \textit{for any } Y_a \in L(\mathcal{H}, \mathbb{R}_+) \textit{ and } h \in \mathcal{H}$$

**Definition 1.2** (FIFO principle)**.** *An arc travel time model is said to obey the FIFO principle if the map $h \mapsto h + t_a[Y_a](h)$ is increasing for any $Y_a \in L(\mathcal{H}, \mathbb{R})$.*

Finally let us introduce the concept of dynamic transport networks, which is the essential component of our modelling approach of the transport supply.

**Definition 1.3** (Dynamic transport network)**.**

- *A dynamic transport network is a triple* $G = (N, A, \boldsymbol{T}_A)$ *where* $(N, A)$ *is a directed graph and* $\boldsymbol{T}_A = (t_a)_{a \in A}$ *are the associated arc travel time models.*

- *A dynamic transport network state is a triple* $(N, A, \boldsymbol{H}_A)$ *where* $(N, A)$ *is a directed graph and* $\boldsymbol{H}_A = (H_a)_{a \in A}$ *are the associated arc exit time functions.*

**Notes.** The notations presented here are adapted from the ones used in LADTA, which were in turn a dynamic adaptation of the one introduced by (Sheffi, 1985). The use of arc exit time functions for dynamic network models dates back to the origin of dynamic network modelling and was already adopted in Merchant and Nemhauser's (1978) model.

# 3 The continuous dynamic network loading problem

This section is dedicated to the *continuous Dynamic Network Loading Problem* (DNLP). In the DNLP, given time-varying route flows on a transport network, one aims to find arc volumes, arc travel times, and route travel times over a finite time period. The term "loading" originally comes from static user equilibrium literature and initially designated the algorithmic operation of computing the arc flows from the flows assigned on the routes of the network.

The problem may be considered as a subproblem of the dynamic user equilibrium problem, as it allows to build the transport supply (here the route travel times) from a transport demand (here the route flows). We will focus on a specific network model, with a single user class, a volume-delay travel time model and no intersection model. However, most of the results presented here can be extended to other travel time models.

**Volume-delay travel time models** On each arc $a$ of the transport network, the arc travel time of a user arriving on $a$ at instant $h$ is solely

determined by the total traffic volume still on $a$ at $h$. Thus for each arc $a$ a volume-delay function $f_a$ taking a traffic volume as input, and returning a travel time. Thus, denoting $S_a : h \mapsto S_a(h)$ the time-varying volume on arc $a$, the travel time for a vehicle entering $a$ at $h$ is given by $f_a\big(S_a(h)\big)$.

The section presents an analytical formulation of the problem (Originally presented in Wu, Chen and Florian, 1998), the existing results on existence and uniqueness (From Xu, Wu, Florian and Zhu, 1999), and exposes a few existing computation procedures. The text is structured along those lines.

## 3.1 Formulation of the DNLP

Informally, the DNLP consists in determining arc volumes and arc travel times given the route flow vector $\boldsymbol{X}_R$. In order to precisely state the DNLP, it is necessary to expose the fundamental equations of the network flowing model. This set of equations is exposed below. It implicitly assumes that the travel time functions resulting from the loading are FIFO *i.e.* that the maps $h \mapsto h + \tau_a(h)$ are increasing. Since the volume-delay travel time model does not respect the FIFO principles, this needs to be enforced by some ways. This formulation is essentially the one presented in (Wu et al., 1998) with the notations introduced in the previous section.

Consider a dynamic transport network $(A, N, T)$ with $T = (t_a)_{a\in A}$. Let us introduce the following definitional equations.

**Cumulated flow on arc** $a$ The cumulated flow on an arc is the sum of the route cumulated flows on each route traversing this arc, translated by the corresponding travel times. The travel times on each arc are assumed to be FIFO[5]. Under this assumption the vehicles following route $r$ have entered on the arc $a$ before $h$ if and only if they departed before $H^{-1}_{r\preceq a}(h)$. Formally[6]:

$$X_a(h) = \sum_{r:a\in r} X_r \circ H^{-1}_{r\prec a}(h) \quad \forall a \tag{1.5}$$

[5]In some cases no loading solutions will satisfies this assumptions and thus the question of the consistency of the FIFO behaviour with the other assumptions is relevant. More comments on this problem will be given latter.

[6]Note that thanks to the FIFO assumption the functions $H^{-1}_{r\prec a}$ are well defined.

**Evolution of the traffic volume on arc** $a$ The evolution of the volume of traffic on arc $a$ is then described by:

$$S_a(h) = Y_a(h) - Y_a \circ H_a^{-1}(h) \quad \forall h, a \tag{1.6}$$

**Arc Travel time model** In the simple flowing model considered in this section, the arc traversal time of a vehicle stems straighforwardly from the traffic volume on this arc at the time of arrival in the node. Formally:

$$\tau_a(h) = f_a\big(S_a(h)\big) \text{ and } H_a(h) = h + \tau_a(h) \quad \forall h, a \tag{1.7}$$

**Route travel time and route exit time computation** The route travel time on route $r$ for a departure at $h$ is defined as the summation of arc travel times along $r$. Each arc travel time function is evaluated at the arrival time on that arc or equivalently at the departure time of the preceding arc (since no waiting is allowed on nodes).

$$\text{for any } r = a_1, \ldots, a_n\text{: } H_r(h) = H_{a_n} \circ \ldots \circ H_{a_1}(h) : \tag{1.8}$$

and:

$$\tau_r(h) = H_r(h) - h \tag{1.9}$$

The continuous dynamic network loading problem can now be formally stated:

**Definition 1.4** (The continuous dynamic network loading problem)**.** *Assume a transport network represented by a graph* $(N, A, \boldsymbol{T}_A)$ *where each arc travel time model* $t_a$ *is a volume-delay travel time model and denote* $f_a$ *the corresponding volume-delay function. Given a route cumulated flow vector* $\boldsymbol{X}_R$*, find an arc cumulated flow vector* $\boldsymbol{Y}_A$*, together with the associated* $(H_r)_{r\in R}$*,* $(\tau_r)_{r\in R}$*,* $(\tau_a)_{a\in A}$ *and* $(S_a)_{a\in A}$*, satisfying Equations (1.5-1.9) on* $\mathcal{H}$ *with initial conditions* $Y_a(h_m) = S_a(h_m) = 0$ *for all* $a \in A$*.*

## 3.2 Comments about the formulation of the DNLP

**Cyclic dependency.** Equations (1.5-1.9) describe a cyclic sequence of relationships between the state variables of the problem. Given the route exit time functions $(H_r)_{r\in R}$, the arc cumulated flow vector can be deduced (Equation (1.5)), and the arc volumes as well as the travel time then straightforwardly derive from Equations (1.6) and (1.9). Finally Equation (1.8) yields $(H_r)_{r\in R}$. The dependency circle is represented in Figure 1.3.

This structure yields the question of the adequate output variables to consider. In Definition 1.4, the problem is formulated in terms of arc cumulated flows and the other quantities are seen as deriving from them. Yet the problem might be similarly formulated in terms of route travel time or arc travel times functions, route exit time or arc exit time functions, or even in terms of arc volumes.

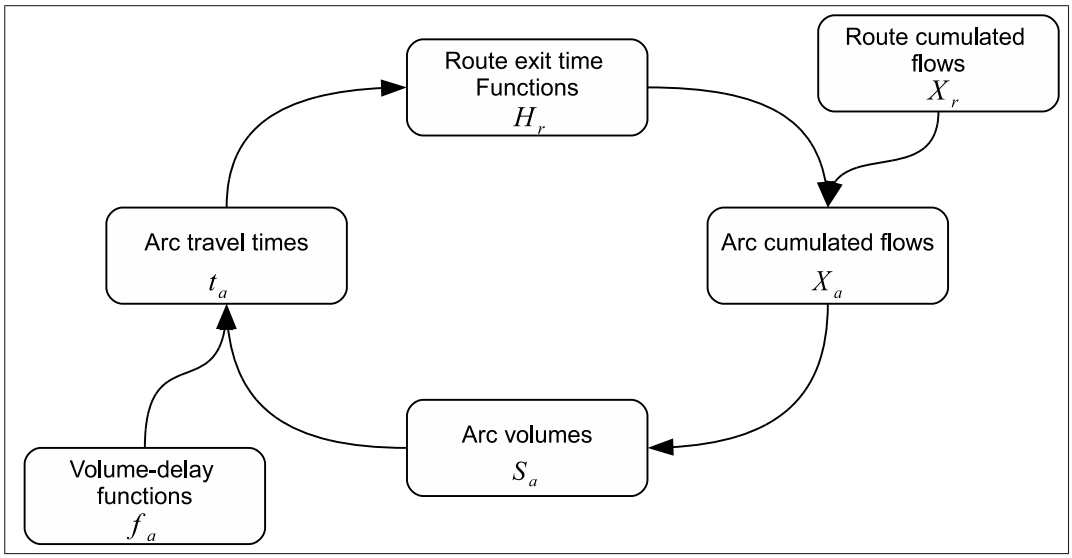



Figure 1.3: Relationship structure in the dynamic network loading problem adapted from (Leurent, 2003*a*)

**FIFO assumption.** The network loading model presented in subsection 3.1 is only valid if the travel time functions corresponding to the solution of the DNLP are FIFO *i.e.* for all $a \in A$ the map $h \mapsto h + \tau_a(h)$ is increasing. In the general case volume-delay model does not respect the FIFO principle so it might not always be the case. As already mentioned there are two ways for imposing the FIFO principle on a volume-delay model: either by restricting to linear volume-delay function or by restricting the set of admissible arc cumulated flows. Both methods strongly limit the scope of the model.

## 3.3 Formal properties of the continuous dynamic network loading problem

The following theorem is due to Wu et al. (1998).

**Theorem 1.5** (Adapted from Wu et al. (1998))**.** *Assume the volume-delay functions are strictly positive, non-decreasing and continuously differentiable functions. Then if the arc FIFO condition is guaranteed the DNLP has a unique solution.*

Theorem 1.5 thus states the conditions under which the problem is well-posed. Its main flaw is the FIFO conditions – as mentioned earlier it can not be guaranteed without strongly restricting the scope of the model. Nevertheless it is to our knowledge the only existence theorem available for the problem and its proof is constructive thus providing interesting insights into the problem structure. Although we are not going to provide the details of the proof, the general ideal is sketched just below.

*Idea of the proof.* The proof works with a reformulation of the problem that introduces a new quantity $Y_{a,r} := X_r \circ H_{r \prec a}$ defined for all $a$ and $r$. $Y_{a,r}$ physical interpretation is the cumulated flow of traffic entering on arc $a$ while following route $r$. it is sensible to replace Equation (1.5) by the two following equations:

$$Y_a(h) = \sum_{r:a \in r} Y_{a,r}(h) \tag{1.10}$$

$$\text{for any } a \in r = a_1, \ldots, a_n\text{: } Y_{a,r}(h) = \begin{cases} X_r(h) & \text{if } a = a_1 \\ Y_{a_{i-1},r} \circ H^{-1}_{a_{i-1}}(h) & \text{if } a = a_i \neq a_1 \\ 0 & \text{otherwise.} \end{cases} \tag{1.11}$$

It is possible (and relatively easy) to show that solving the DNLP as presented in Definition 1.4 is equivalent to find a set of quantities $\boldsymbol{Y}_A$, $(Y_{a,r})_{a \in A, r \in R}$, $(H_r)_{r \in R}$, $(\tau_r)_{r \in R}$, $(\tau_a)_{a \in A}$ and $(S_a)_{a \in A}$, satisfying Equations (1.10), (1.11) and (1.6-1.9) on $\mathcal{H}$, with initial conditions $Y_a(h_m) = S_a(h_m) = 0$ for all $a \in A$.

The proof then proceeds by induction.

**Base case.** Denote $\tau_m := \min_a f_a(0)$. The quantity $\tau_m$ represents the minimum time to traverse an arc of the network. For any $r = a_1 \ldots a_n$, set

$$Y_{a,r}(h) := \begin{cases} X_r(h) & \text{if } a = a_1 \text{ and} \\ 0 & \text{otherwise,} \end{cases}$$

for any $h \in [h_m, h_m + \tau_m]$. Then use Equations (1.10), (1.6), and (1.9) to derive the quantities $Y_a$, $S_a$ and $\tau_a$ for all $a$ on the interval $[h_m, h_m + \tau_m]$.

**Induction step.** Assume $Y_{a,r}$, $S_a$, $\tau_a$ are known until $h_k$ and set $\tau_k := \min_a f_a\big(S_a(h_k)\big)$ (note that $\tau_k \geq \tau_m$). Set $Y_{a,r}$ on $[h_k, h_k + \tau_k]$ according to (1.11). Derive the quantities $Y_a$, $S_a$ and $\tau_a$ accordingly.

It is then necessary to show that the induction terminates in a finite number of steps *i.e.* that all the traffic will have exited the network after a certain time. It is easy to check that the arc cumulated flow vector $\boldsymbol{Y}_a$ hereby built is a solution to the DNLP. The quantity $H_r$ and $\tau_r$ can be derived straightforwardly from Equations (1.8) and (1.9). The uniqueness property requires a few more mathematical precautions that can be found in the original article. □

An important consequence of this theorem is that, given a specific dynamic network $G = (A, N, T)$, one can build a set of route travel time models that associate to a route flow vector $\boldsymbol{X}_R$ the corresponding route travel time vectors. In the following, we will often denote those functions $t_r[\boldsymbol{X}_R]$.

## 3.4 Solution methods

**Mathematical programming.** To our knowledge the first attempt to solve the DNLP is a mathematical programming approach due to Wu et al. (1998). In their method the output variables are the route exit time functions $H_r$. The global idea is to note that given a candidate vector of route exit time functions $\hat{\boldsymbol{H}}_R := (\hat{H}_r)_{r\in R}$, one can compute a second sequence of route exit time functions $\tilde{\boldsymbol{H}}_R := (\tilde{H}_r)_{r\in R}$ from Equations (1.5-1.9). If $\hat{\boldsymbol{H}}_R$ and $\tilde{\boldsymbol{H}}_R$ are equal then $\hat{\boldsymbol{H}}_R$ yields a solution to the DNLP.

The authors propose to formalize this idea under the following minimization program:

$$\min_{\hat{H}_r^{-1}} \sum_{r\in R} ||\hat{H}_r^{-1} \circ \tilde{H}_r - \mathrm{id}_{\mathcal{H}}||_2 \tag{1.12}$$

subject to:

$$
\begin{aligned}
X_a(h) &= \sum_{r:a\in r} X_r \circ \hat{H}^{-1}_{r\prec a}(h), \qquad \text{for all } a \in A \\
S_a(h) &= Y_a(h) - Y_a \circ \hat{H}^{-1}_a(h), \quad \text{for all } a \in A \\
\tilde{H}_{r\prec a}(h) &= h + \sum_{a'\in\{r\prec a\}} t_{a'}\big(H_{r\prec a'}(h)\big), \quad \text{for all } r \in R
\end{aligned}
$$

(1.12) is a non-convex infinite dimensional minimization program whose solutions are solutions to the DNLP. Wu et al. (1998) propose an approximation of the program by discretizing the set of departure times $\mathcal{H}$ and by considering polynomial approximations of the inverse exit time functions $\hat{H}^{-1}_r$. The program is then expressed under a GAMS implementation and solved using MINOS solvers. GAMS (for General Algebraic Modelling System) is computer langage designed to represent and solve large and complex mathematical programming problems. MINOS solvers are state of the art solvers for non-linear and non-smooth optimization problems (Brooke, Kenderick, Meeraus and Release, 1996).

The results are not very convincing. For small size networks (approx. 10 arcs), it is possible to solve the problem within a reasonable computing time and efficiency, but for larger networks it quickly becomes untractable. This is not very surprising since global (*i.e.* in this context non-convex) optimization is a difficult topic especially when dealing with some much dimensions and when the optimization program has no specific mathematical property.

**Chronological computation.** An alternative approach has been proposed by Xu et al. (1999) under the name of the DYNALOAD algorithm and then improved by Rubio-Ardanaz, Wu and Florian (2003). These methods, which may be considered as event-based simulations, represent a major improvement over the mathematical programming approach exposed previously.

This method requires a specific assumption: the volume-delay functions $f_a$ are assumed to be strictly positive, so the travel times on an arc of the network is never 0. The algorithm is based on the constructive proof of Theorem 1.5. It is made explicit below (Algorithm 1.1). Numerical benchmarks have shown that the algorithm can be used efficiently on large size networks (a few thousands arcs). Yet the number of iterations is possibly very high and strongly depends on the network's topology: when short (in terms of travel time) arcs exist the number of iterations naturally increases.

**Algorithm 1.1** `Dynaload`$(\boldsymbol{X}_R,(f_a)_{a\in A},\mathcal{H},\bar{\mathcal{H}})$

**Inputs:** $\mathcal{H} = [h_m, h_M]$, the set of departure times
$\bar{\mathcal{H}} = [h_m, \bar{h}_M]$, the simulation period
$\boldsymbol{X}_R = (X_r)_{r\in R}$, a route cumulated flow vector
**Outputs:** An arc cumulated flow vector $(Y_a)_{a\in A}$
**Initialize** $h_k := h_m + \min_a f_a(0)$
$Y_{a,r}(h) := X_r(h)$ for all $(a,r) : r = a \ldots$
$S_a(h) := 0$ for $h \in [h_m, h_k]$ and all $a \in A$
**While** $h_k < \bar{h}_M$
**Compute** $S_a$ on $[h_m, h_k]$ from Eqn (1.9) for all $a \in A$
**Set** $h_{k+1} := h_k + \min_a f_a(S_a)$
**Derive** $Y_{a,r}$ on $[h_k, h_{k+1}]$ from Eqn (1.11) for all $a \in A$
**Set** $k := k + 1$
**End While**

**Fixed point approaches.** Finally let us introduce a last category of algorithms that we termed as fixed point procedures. To our knowledge it was initially proposed by Chabini (2001) but since then it has widely been embedded in dynamic traffic assignment algorithms (see next chapter). Essentially the method recognizes the fixed point structure of the DNLP. Indeed one need to find arc travel time that are consistent with arc cumulated flows (*i.e.* $(t_a)_{a\in A}$ that yield to $\boldsymbol{Y}_a$ by Equations (1.6) and (1.7)). The procedure is an adaptation of the method of successive averages to this case.

**Algorithm 1.2** `FixedPointLoading`$(\boldsymbol{X}_R,(f_a)_{a\in A},\mathcal{H})$

**Inputs:** $\mathcal{H} = [h_m, h_M]$, the set of departure times
$\boldsymbol{X}_R = (X_r)_{r\in R}$, a route cumulated flow vector
$(f_a)_{a\in A}$, the arc volume-delay functions
**Outputs:** An arc cumulated flow vector $\boldsymbol{Y}_A = (Y_a)_{a\in A}$
**Parameter:** $w_k$ a decreasing sequence from 1 to 0.
**Initialize** $\tau_a^{[0]}(h) := f_a(0)$ for all $a$ and $h$ and $k := 0$
**Do**
**Set** $k := k + 1$
**Compute** the cumulated flows $Z_a$ from $\tau_a^{[k]}(h)$ and Eqn (1.8) and (1.5)
**Set** $Y_a^{[k]} := w_k.Y_a^{[k-1]} + (1 - w_k).Z_a$ for all $a \in A$
**Compute** $\tau_a^{[k]}$ from $Y_a^{[k]}$ and Eqn (1.6) and (1.7)
**Until** $\boldsymbol{Y}_A^{[k]}$ **satisfies a certain criterion**

Note that this algorithm can be applied to virtually any arc travel time models. Algorithm 1.2 has been tested in details on medium size networks (*i.e.* a few hundreds arcs) in (Chabini, 2001) and was shown to converge well although at a slower rate than chronological methods. Now its integration by Bellei et al. (2005) in a dynamic traffic assignment procedure showed it was very useful to quickly compute approximate solutions to the DNLP.

# 4 Dynamic shortest and least cost routes

Shortest route problems are among the most studied problems in graph theory and their resolution is known considered as routine (Deo and Pang, 1984). An important number of extensions have been considered, for instance shortest routes within a time windows (Desrosiers, Soumis and Desrochers, 1984)

or shortest route with non-linear value of time. The great majority of this extensions are dealing with static networks that have fixed travel times and fixed costs. The present section focus on dynamic shortest and least cost route problems. There are two understandings for a dynamic shortest route problem. In the first, one must recompute shortest routes due to frequent, instantaneous, and *unpredictable* changes in network data. This is essentially a reoptimization problem, involving the resolution of a sequence of closely-related static shortest route problems. The second understanding is the time-dependent shortest route problem, in which network characteristics change with time in a *predictable* fashion. This is the version usually studied in transport science and the topic of this section.

The problem was initially introduced by Cooke and Halsey (1966) with the costs limited to the travel times and under a discrete time formulation. A well known result from Dreyfus (1969) is that under the assumptions that the travel times follows a First In First Out discipline, the question of finding the shortest route in this case boils down to a static shortest route by extending the network through time. Recently a renewed interest in those algorithms has appeared due to its applications for transport forecasting (or more exactly dynamic user-equilibrium computation) and in intelligent transport system research.

The section is divided in three subsections. First we expose in details the shortest route problem on FIFO networks, then a review of the known solution algorithm is proposed and finally some extensions are considered.

## 4.1 Dynamic shortest route on a FIFO network: statement and properties

**Notations and comments.** The notations are the one exposed in Section 2. In a dynamic shortest route problem, we consider a dynamic transport network state $G = (N, A, \boldsymbol{H}_A)$ with $\boldsymbol{H}_A = (H_a)_{a_i nA}$. The quantity $H_a(h)$, assumed to be such that $H_a(h) > h$ gives the exit time of $a$ if one enters at time $h$. Note that papers in the literature tend to work with arc travel time functions rather than exit time functions. Obviously this is equivalent but exit time functions leads to simpler formulations. In this section the network is said to be a *FIFO network* if all exit time functions $H_a$ are FIFO in the sense of the previous section *i.e.* $H_a$ is non-decreasing.

There are two essential variants of the shortest route problem, whether

*time is represented continuously* or *not*. In discrete-time settings, the arc exit time functions $H_a$ can be identified to integer valued functions of an integer argument. In continuous-time settings, $H_a$ are real-valued functions defined on a real set. Most of the results presented here are valid for the two settings.

In this first subsection, we study in detail the shortest route problem in FIFO networks. We will see that this assumption leads to very rich theoretical properties that are no longer valid in general networks.

**Statement of the shortest route problems in FIFO networks.** Recall that $R_{od}$ denotes the set of routes serving the origin destination pair $od$. We introduce the two basic quantities that one might want to find in a dynamic shortest route.

$$H_{od}(h) := \min_{r \in R_{od}} H_r(h) \tag{1.13}$$

$$\bar{H}_{od}(\bar{h}) := \max_{r \in R_{od}} H_r^{-1}(\bar{h}) \tag{1.14}$$

$H_{od}$ is the *earliest arrival time* function while $\bar{H}_{od}(\bar{h})$ is the *latest departure time* function.

The simplest variants of the dynamic shortest route problem are to find $H_{od}(h)$ or $\bar{H}_{od}(\bar{h})$ for a fixed OD pair $od$ and a fixed departure time $h$ or a fixed arrival time $\bar{h}$. Many other variants are possible whether one wishes to consider a range of origins, destinations or departure times. Table 1.1 summarizes the variants using a wildcard notation introduced by Dean (2004*b*). For instance the problem of computing $H_{o*}(*)$ is the problem of computing $H_{od}(h)$ for all $d$ and $h$. The wildcard notation allows to distinguish 16 variants of the problem, but at the end only two fundamental problems need to be addressed.

This reduction requires a time-reversal transformation to change earliest arrival time problems into latest departure time problems. The operation simply consists in reversing the direction of the arcs and inverting the associated exit time functions.

Note that we choose the length of the routes as output for the dynamic shortest route problem rather than the routes themselves. In practice the structure of the algorithms used in practice to solve the dynamic shortest route problem always allows to compute explicitly the shortest routes.

It is important to understand that all the problems of Table 1.1 are of interest for dynamic user equilibrium computation. Early arrival problems are

required when considering dynamic user equilibrium models with only route choice, while late departure problems needs to be solved for models combining route and departure time choice. When considering morning commute, Chabini (1998) argues that as users tend to converge from an important number of origins to a few destinations, all origin to one destination shortest route problems are important for transport modellers. The same argument stands for evening commutes and all destinations to one origin problems.

| **Desired output** | **Method of computation** |
|---|---|
| $H_{o*}(h)$, $H_{o*}(*)$ | The two fundamental problems. All other variants can be expressed with these two. |
| $H_{od}(h)$, $H_{od}(*)$ | As in the static case, single-origin, single destination problems are as difficult as multiple-origin multiple destinations ones. Therefore they can be solved by computing the more general $H_{o*}(h)$, $H_{o*}(*)$. |
| $H_{**}(h)$, $H_{**}(*)$ | Perform a computation of $H_{o*}(h)$ or of $H_{o*}(*)$ for each origin. |
| $H_{*d}(h)$ | This problem is no easier than computing $H_{**}(*)$. |
| $\bar{H}_{od}(*)$ $\bar{H}_{o*}(*)$ $\bar{H}_{*d}(*)$ $\bar{H}_{**}(*)$ | Find and invert the corresponding earliest arrival time function. Alternatively perform a time-reversal transformation of the network. |
| $\bar{H}_{o*}(h)$ | This problem is no easier than computing $\bar{H}_{**}(*)$. |
| $\bar{H}_{*d}(\bar{h})$ | Perform a time-reversal transformation of the network and compute $H_{o*}(h)$. |
| $H_{*d}(*)$ | Perform a time-reversal transformation of the network and compute $\bar{H}_{o*}(*)$. |
| $\bar{H}_{od}(\bar{h})$ | Solve the more general problem $\bar{H}_{*d}(\bar{h})$. |
| $\bar{H}_{**}(\bar{h})$ | Perform a computation of $\bar{H}_{o*}(\bar{h})$ for each origin. |

Table 1.1: Reduction of the dynamic shortest route problem to two fundamental variants (adapted from Dean, 2004*b*)

**Shortest route problems' properties for FIFO networks and optimality condition.** For early arrival time problems, the following proper-

ties of FIFO networks have been formally established by Kaufman and Smith (1993) for the discrete case and by Dean (2004*b*) for the continuous case.

**Proposition 1.6** (Shortest route problems' properties for FIFO networks)**.** *The following properties stand for any shortest route in a FIFO network:*

**No-waiting** *In a FIFO network, waiting at nodes is never beneficial i.e. it never reduces the arrival time at destination.*

**Acyclicity** *In a FIFO network, one may always find shortest routes which are acyclic.*

**Route consistency** *In a FIFO network, one may always find shortest routes whose subroutes are also shortest routes.*

*Idea of the proof of Proposition 1.6.* See (Kaufman and Smith, 1993) for a detail proof of the discrete case and (Dean, 2004*b*) for the continuous one.

The no-waiting property follows directly from the fact that arc arrival time functions are non-decreasing. Acyclicity is a consequence of the no-waiting property: assume a shortest route with cycle and replace the cycle with the corresponding waiting time. Then, either the route is not a shortest route of the travel time (a contradiction from the no-waiting property) or the cycle has null travel time. □

These properties highlight the interest of the FIFO assumption for dynamic shortest route problems. The no waiting and acyclicity properties guarantee that the problem, as we stated it, is consistent. The route consistency property allows to state the optimality principle above, which is the base of most algorithmic approaches for dynamic shortest route computation. It is important to note that this property is not true if the FIFO assumption is not valid as depicted in Figure 1.4.

**Proposition 1.7** (Optimality condition)**.** *The following condition is necessary and sufficient for the dynamic shortest route problems $H_{o\star}(\star)$ and $H_{o\star}(h)$ on FIFO networks. For $H_{o\star}(h)$ it must hold for a fixed $h$, while for $H_{o\star}(\star)$ for any instant $h$ in $\mathcal{H}$.*

$$H_{om}(h) = \begin{cases} \min\limits_{n \in N_-(m)} H_a\big(H_{on}(h)\big) & \text{if } m \neq o \\ 0 & \text{if } m = o \end{cases} \tag{1.15}$$

*where $N_-(m)$ denote the set of the parents nodes of $m$.*

The optimality condition was first proposed by Cooke and Halsey (1966) without a formal proof, and then properly established by Orda and Rom (1991). The proof straightforwardly stems from the route consistency property.

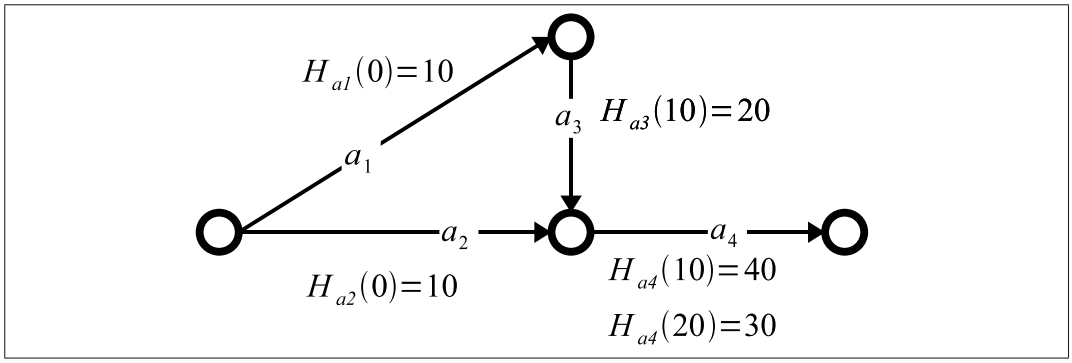



Figure 1.4: Route inconsistency in non-FIFO networks

## 4.2 Algorithms

In this subsection we describe the main algorithm addressing the shortest route problem in FIFO networks. We denote $n = |N|$ the number of nodes, and $m = |A|$ the number of arcs.

**Computing $H_{o\star}(h)$.** The problem of computing $H_{o\star}(h)$ was the first to be addressed from a computational perspective. Cooke and Halsey (1966) consider the problem by viewing a discrete dynamic network as a static network by using a time-space expansion of the network. Later Dreyfus (1969) showed the problem could be treated in a similar manner as the static case. In particular the classic label setting (Dijkstra's) and label correcting (Ford Bellman's) algorithm can be straightforwardly extended to cope with dynamic network and their resulting complexity is the same than for the corresponding static network. The running time are thus in $O(m + n\ln(n))$ for label-setting algorithm and $O(mn)$ for label-correcting ones. The pseudocode for both algorithms is given in Algorithms 1.3 and 1.4.

**Computing $H_{o\star}(h)$ - the naïve approach.** Let us now turn to the problem of computing the earliest arrival for all departure times simultaneously.

**Algorithm 1.3** `DynamicDijsktra`$(o, h, (H_a)_{a\in A})$

**Inputs:** $(H_a)_{a\in A}$, arc arrival time functions
$o$, the origin node
$h$, an instant
**Outputs:** $(H_{od}(h))_{d\in N}$
**Initialize** $S := N$
**Foreach** $n \in N$: $H_{on}(h) := +\infty$
$H_{oo}(h) := h$
**While** $S \neq \emptyset$
**Find and remove** $n \in S$ minimizing $H_{on}(h)$
**Foreach** $a = (n_1, n_2) \in A$
$H_{on_2}(h) := \min\{H_{on_2}(h), H_a\big(H_{on_1}(h)\big)\}$
**End While**

**Algorithm 1.4** `DynamicFordBellman`$(o, h, (H_a)_{a\in A})$

**Inputs:** $(H_a)_{a\in A}$, arc arrival time functions
$o$, the origin node
$h$, an instant
**Outputs:** $(H_{od}(h))_{d\in N}$
**Initialize** $Q := \{o\}$
**Foreach** $n \in N\{o\}$: $H_{on}(h) := +\infty$
$H_{oo}(h) := h$
**While** $Q \neq \emptyset$
**Pop** $n \in Q$
**Foreach** $a = (n_1, n_2) \in A$
**If** $H_a(h) \neq \min\{H_{oj}(h), H_a\big(H_{oj}(h)\big)\}$ **Then:**
$H_{aj}(h) := \min\{H_{oj}(h), H_a\big(H_{oj}(h)\big)\}$
$Q := Q \cup \{a\}$
**End While**

Naturally it can be solved by using repeatedly the previous algorithmic approaches. In a discrete setting, this allows to solve exactly the problem, while in a continuous setting only an approximation is obtained. When $T$ is the number of time steps, the running time is in $O\big(T(m + n \ln n)\big)$.

**Computing $H_{o\star}(\star)$ - Label correcting algorithms.** One may also notice that the label correcting algorithm (Algorithm 1.4) can be straightforwardly extended to the problem of computing $H_{o\star}(\star)$ rather than $H_{o\star}(h)$. This is achieved by updating functions $h \mapsto H_{on}(h)$ rather than scalars $H_{on}(h)$. With this modification one simultaneously computes $H_{o\star}(h)$ for all departure times $h$. This algorithm was proposed in discrete time by Ziliaskopoulos and Mahmassani (1993) for the symmetric problem of computing $\bar{H}_{\star d}(\star)$ and by Orda and Rom (1991) for continuous time settings.

In discrete time setting, the running time is $O(mnT)$. In the continuous time case, computation time depends on the representation of exit time functions and on the computation of the basic operations on the functions (*i.e.* addition, minimum and comparison). If they are implemented as piecewise linear continuous linear functions, then this operations depends on the average number of linear pieces in the functions $H_{on}(h)$. Denoting them $P$ the running time has a complexity of $O(mnP)$. As there is no way of guessing *a priori* $P$, the running time of the algorithm is impossible to predict.

**Computing $H_{o\star}(\star)$ - Label setting algorithms.** The label setting algorithm (Algorithm 1.4) can not be extended directly to treat simultaneously all the departure instants. However, label setting algorithms, in the sense that they compute the solutions in small pieces without updating them during the process. Contrary to the previous cases no unified algorithm has been established for both discrete and continuous settings.

In discrete settings, label-setting algorithms have been initially introduced by Cai, Kloks and Wong (1997) and Chabini (1998). They rely on the fact that the time-expansion of the network is acyclic as soon as the arc travel time are strictly positive (*i.e.* for all arc $H_a(h) > h$). Now in acyclic graph shortest routes might be computed in linear time with respect to the number of arcs, once a topological ordering has been computed. The static shortest route problem corresponding to the dynamic one is thus easy to compute. This is even more the case as the topological ordering in the time-expansion of a dynamic network is straightforward to compute, it may be obtained by

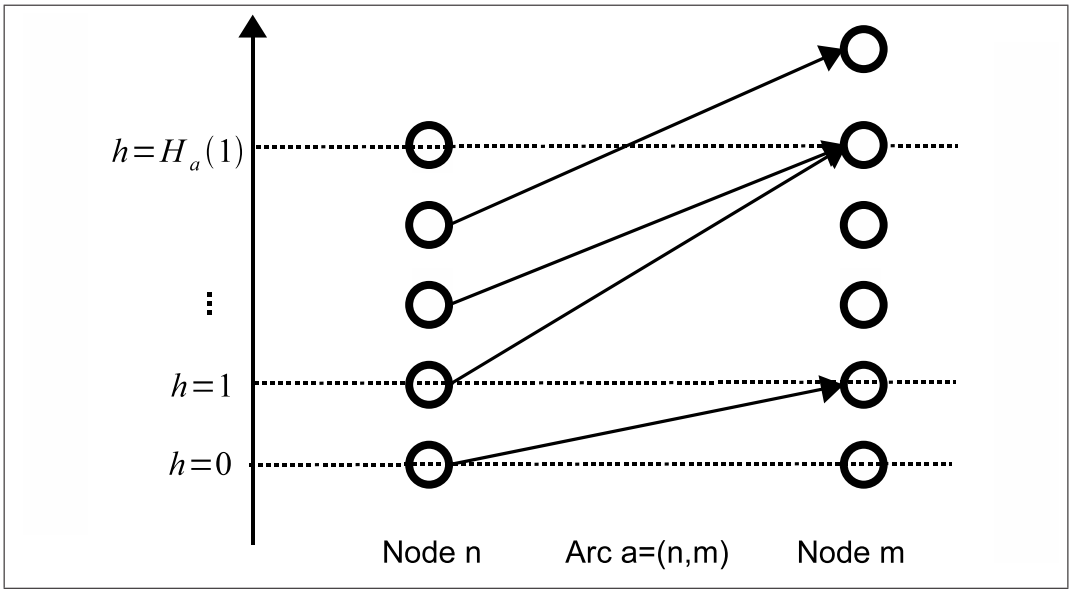



Figure 1.5: time-expansion of a dynamic network

enumerating the nodes in the time-expanded network in chronological order. This approach has a running time in $O\big(T(n+m)\big)$ which matches with the lower bound on the problem complexity. The pseudo-code is presented in Algorithm 1.5, where the problem of computing $\bar{H}_{o\star}(\star)$ is first solved and then $H_{o\star}(\star)$ is deduced. It is noteworthy that the time expanded network is not built explicitly when running the algorithm.

In continuous setting, a label-setting algorithm has been proposed by Dean (1999) for piecewise linear inputs. the algorithms essentially consists in a single chronological scan through time. The resulting complexity is in $O(mP \ln n + n)$ where $P$ is the average number of linear pieces in an output function.

The question of the difference in running times of the discrete and continuous time algorithms is an interesting one. The algorithmic complexity in continuous setting depends on the arc exit time functions by the mean of $P$, while in the discrete setting only the topological characteristics of the network (and the time step) plays a role. One may wonder if the use of the discrete-time algorithm might not be adequate to obtain a good approximation of the continuous problem. The question to answer is then the choice of an appropriate time step to discretize the set of departure times. Ideally one would choose a discretization such that $T$ is significatively larger than

$P$, so that the discrete time solution correctly approximates the exact one. This rises two comments: first there is no way to *a priori* estimate $P$ so in practice $T$ should be chosen to be very large in order to deal with worst case scenarios; second assuming $m = O(n)$ then the complexities of the two algorithms only differ from a $\ln n$ factor. In practice the discrete approximation of dynamic shortest route is of little interest regarding the loss of accuracy it causes.

**Algorithm 1.5** `AllDepartureTimeDynamicLabelSetting`$(o, \mathcal{H}, (H_a)_{a\in A})$, N

**Inputs:** $(H_a)_{a\in A}$, arc arrival time functions
$o$, the origin node
$\mathcal{H} = 1, \ldots, T$, the set of departure times
**Outputs:** $(H_{od}(\star))_{d\in N}$
**Initialize** $Q := \{o\}$
**Foreach** $n \in N\{o\}$ and $h \in \mathcal{H}$: $H_{on}(h) := +\infty$
$H_{oo}(h) := h$
**For** $\bar{h} = 1, \ldots, T$
**Foreach** $a = (n_1, n_2) \in A$
**If** $\bar{H}_{on_2}(h) \leq T$ then
$\bar{H}_{on_2}(\bar{H}_a(\bar{h})) := \max\{\bar{H}_{on_2}(\bar{H}_a(\bar{h})), \bar{H}_{on_1}(h)\}$
**End For**
**Set** $H_{on} := \bar{H}_{on}^{-1}$ for all $n$

## 4.3 Dynamic least cost route problem

There are at least three ways to generalize the previous shortest route problem in FIFO networks:

- *Non-FIFO travel time functions* may be considered.

- *Least cost route* rather than shortest route problems can be introduced by associating time-varying costs functions to each arc.

- *Waiting* might be allowed. In FIFO networks this was not an issue, as waiting is never beneficial. When least cost route rather than shortest route problems are considered, one need to specify waiting costs. Dean (2004*a*) distinguishes three cases: infinite waiting costs (waiting

| | Shortest routes in FIFO networks | | Least cost routes in non-FIFO networks | |
|---|---|---|---|---|
| Algorithm | Discrete time | Continuous time | Discrete time | Continuous time |
| Label correcting | $O(mnT)$ (Ziliaskopoulos et al., 1993) | $O(mnP)$ (Orda and Rom, 1991) | $O(mnT^2)$ (Ziliaskopoulos et al., 2000) | Not polynomial (Orda and Rom, 1991) |
| Repeated $H_{o\star}(h)$ | $O\big(T(m + n\ln n)\big)$ (Dreyfus, 1969) | Not applicable | Not applicable | Not applicable |
| Label setting | $O((m+n)T)$ (Chabini, 1998) | $O(mP\ln n + n)$ (Dean, 1999) | $O((m+n)T)$ (Dean, 2004*a*) | Unknown (forbidden waiting) (Leurent, 2004) |

Table 1.2: Running times of the algorithm for the computation of $H_{o\star}(\star)$. The least cost route running time are for non-FIFO networks with location dependent waiting costs or forbidden waiting.

is forbidden), duration dependent waiting costs (the waiting costs is a function of the waiting time) or location dependent waiting costs (each node is endowed with a functions of the time and its integration on the period of waiting gives the waiting costs). Note that duration dependent waiting costs can be used to encode bounded waiting time constraints.

We call the problems obtained by extending the shortest route problems in FIFO networks the *dynamic least cost route problems*. From a transport modelling perspective, dynamic least cost route problem are especially appealing. Non-FIFO networks allow to represent relevant traffic phenomena, such as overtaking. Second some commonly used travel time model typically generate non-FIFO travel time function (*e.g.* volume-delay functions). Considering least cost route problems rather than shortest route allows to model the trade-off between travel time and monetary cost on tolled networks. But the most promising extension is probably the addition of the various waiting costs. Obviously it can be exploited directly to model some observed travel behaviours. For instance in London subsequent to the congestion charge implementation, some users started to wait for the end of the charged period (at 6pm) before entering the charged zone causing important congestion in the parking spaces nearby. A second important application is the use of the

waiting costs to encode possible activities at stops. This can be achieved by considering negative waiting costs that represent the utility derived from the performance of a specific activity.

From a computational perspective in non-FIFO networks are more complicated to handle as the interesting properties of Proposition 1.6 are no longer valid. For instance a route might be cyclic. Moreover the problem is proved to be NP-complete in discrete setting as a reduction from the knapsack problem (Cai, Sha and Wong, 2007).

In discrete time, the algorithms presented below can be easily adapted to dynamic least cost route problems although with some degradation in computation time. When costs are limited to travel times in a non-FIFO network, Algorithm 1.5 can be applied with no modification. Table 1.2 summarizes the known complexity results for dynamic least cost route problems assuming that waiting costs are duration dependent. Note that the algorithms based on the time-expansion of the dynamic networks still remain the most efficient way to deal with the problem.

A complex issue is the one of taking account of the waiting costs. Location dependent costs can trivially be incorporated in a discrete time network by adding fictionnous arcs with constant unit travel times and adequate costs functions. However duration dependent costs are much more complicated to deal with. A good survey of the algorithmic techniques to deal with this case is presented in (Dean, 2004*b*).

In continuous time setting the general problem is complex and exhibits some surprising features. Orda and Rom (1990) showed that no finite optimal route exist in some networks. Now by taking into account infinite routes, one could guarantee the existence of an optimal route (Orda and Rom, 1991). Moreover with some reasonable assumptions on the arc travel time and costs functions, the optimal route is finite. The existing algorithms are rare. Leurent (2006) presents a general theory, the so-called dynamic network theory, allowing to treat very general shortest path problem with a label setting or label correcting algorithms. Notably least cost route problems in non FIFO networks falls under that category. Additionally the theory allows to deal easily with constrained least cost route problems. Orda and Rom (1991) designed a label correcting algorithms for least cost route problems with location dependent waiting costs, again with piecewise linear exit time and costs functions. To the author's knowledge the design of an algorithm for location dependent costs is still an open question.

## 4.4 Other issues of interest

**Parallelization.** As transport applications require to perform dynamic least cost route problem on large networks, parallelization strategies have been explored in the literature. Chabini and Ganugapati (2002) present a parallelization method for their label-setting algorithms using a network decomposition technique. Ziliaskopoulos, Kotzinos and Mahmassani (1997) introduce parallel designs for shortest route problems in non FIFO networks. The most efficient is based on a destination decomposition technique. In both cases, significant speed up of the equivalent sequential algorithm is achieved.

**Zero travel time.** Recall that we assume that $H_a(h) > h$ *i.e.* that all travel times were strictly positive. Dealing with networks where zero travel times arc might exist is more complicated than it seems. Cai et al. (1997) deals with this issue, but the corresponding running times incur a slight increase in computing time. This is due to the fact that the time expanded network might not be acyclic anymore and that a static shortest route is now required within each "time level".

# Conclusion

This chapter has presented an overview of the main modelling approaches of dynamic transport networks in the context of network equilibrium. It has revealed that road arc congestion has received considerable attention, but that other issues of importance such as congestion in public transport are still poorly taken in account. The two subsequent sections exposed in detail the two main algorithmic problems related to dynamic transport networks.

The dynamic network loading problem has been extensively studied for the volume-delay travel time models, but little attention has been paid to other analytical travel time model. In Appendix D we present an algorithm dedicated to the dynamic network loading problem with bottleneck travel time, thus fulfilling this gap in the literature. The algorithm is strongly inspired by the formalization of the dynamic network loading problem proposed in Chapter 3.

The second problem, the dynamic least cost route, has received considerable attention. Several variants of the problem have been considered, and most of them have real interest from a transport modelling perspective. Al-

though interesting issues remain to be addressed (*e.g.* non additive route costs or bi-criteria problems, both in a time-varying context), to the author's point of view the real challenge is to spread those results among the transport science community. Until now few applied works in transport research use or even quote those works, despite their usefulness. This is especially true for works dealing with the computation of the dynamic user equilibrium.

Chapter 2

# Mathematical formulations for the dynamic user equilibrium

In the previous chapter, the focus was on transport supply. In the present one, we will study transport demand and the precise formulation of the equilibrium between supply and demand. In static models, the seminal works of Wardrop (1952) and Beckmann et al. (1956) set up the reference framework on which state of the art models still draw upon. Following Wardrop and Beckmann, the static user equilibrium principle has been extended to dynamic transport networks. Yet, unlike in the static case, the transport science community is lacking of a unified modelling framework for the dynamic user equilibrium.

There exists a wide variety of alternative equilibrium principles. For example the Boston equilibrium principle[1] proposed by Friesz, Luque, Tobin and Wie (2003) requires that for each instant the instantaneous flows and instantaneous travel costs (*i.e.* costs perceived at the moment of departure) constitute a static user-equilibrium. This situation is sometimes also referred to as a quasi-dynamic traffic equilibrium or reactive user equilibrium. One could also mention all sorts of dynamic stochastic equilibrium principles or the dynamic system optimal equilibrium, although the term equilibrium here is slightly abusive.

One of the most simple and widely used equilibrium principles is the so-

[1] The name Boston equilibrium comes from the authors' experience of driving in Boston where, at that time, an very accurate description of the traffic situation was available by the radio. The generalization of intelligent traffic system caused a renewed interest in those kind of models at the end of the 90s although it is now admitted that they poorly predict reality.

called *dynamic Wardrop principle*, which states that:
*At all instants the journey times on the route actually used are equal and less than those which would be experienced by a single vehicle on any unused route.*
Note that the terminology varies according to the authors and that some might speak of route choice user equilibrium principle. From a behavioural perspective, the dynamic Wardrop principle assumes that the users of the transport network are homogeneous, have perfect information, and that their departure time is exogenous. That said, it can be shown, with the adequate assumptions, that this is equivalent to a "no incentive to change" criterion: given the current pattern of route travel times and given users' route choice, no user would gain by choosing an alternative route. In that sense, a Wardrop equilibrium is closely related to the concept of Nash equilibrium in game theory.

The problem of computing the dynamic Wardrop equilibrium is sometimes refer to as the *dynamic Wardrop assignment problem* and is often presented as a variant of the assignment step of the well known four step model. Now the no incentive criterion can be used to generalize the concept of dynamic Wardrop principle to *dynamic user equilibriums* (DUE). In DUE models, more advanced representations of the transport demand are considered, for instance by allowing users to choose their departure time or considering generalized costs rather than travel times. In this review, we focus on this type of equilibriums *i.e.* it is assumed that the users of the considered transport network have *perfect information* and are acting as *selfish cost-minimizing* agents.

The chapter is structured as follows. The first section presents the dynamic Wardrop assignment problem. The second and third sections present two specific formulations as standard problems, namely variational inequality and fixed point problems. The associated algorithms are reviewed. The last section presents extensions of the dynamic Wardrop assignment to the dynamic user equilibriums that consider more complex representations of the transport demand.

# 1 The dynamic Wardrop assignment problem

We use the notations introduced in the previous chapter. Assume we have a traffic demand represented by an OD matrix $\boldsymbol{X}_{OD} := (X_{od})_{od \in OD}$. The

quantity $X_{od}$ is a cumulated flow defined on the set of departure times $\mathcal{H}$ for the origin-destination pair *od*. An *assignment of the traffic demand* is a route cumulated flow vector $\boldsymbol{X}_R$ such that $\sum_{r \in R_{od}} X_r = X_{od}$. The *route travel times functions* arising from the loading of an assignment of the demand on the network are denoted $t_r[\boldsymbol{X}_R]$ and the map $\boldsymbol{X}_R \mapsto t_r[\boldsymbol{X}_R]$ is a *route travel time model*. Recall that route travel time models $\boldsymbol{X}_R \mapsto t_r[\boldsymbol{X}_R]$ can be computed by solving the dynamic network loading problem and are well-defined according to the existence theorem stated in Chapter 1, Subsection 3.3.

Recall that we model route flows $X_r$ as absolutely continuous functions on a set of instants $\mathcal{H} = [h_m, h_M]$. Consequently $X_r$ admits a derivative almost everywhere that is denoted $x_r$ (see Chapter 1, Section ). As absolutely continuous function on a closed interval is continuous, so are route flow functions.

The dynamic Wardrop assignment is then defined as:

**Definition 2.1** (dynamic Wardrop assignment)**.** *Consider a dynamic transport network described by its route travel time models* $\boldsymbol{t}_R = (t_r)_{r \in R}$ *and an OD matrix* $\boldsymbol{X}_{OD}$*. An assignment* $\boldsymbol{X}_R$ *of* $X_{OD}$ *is a dynamic Wardrop assignment if and only if all* $r, r' \in R_{od}$

$$x_r(h) > 0 \Rightarrow t_r[\boldsymbol{X}_R](h) \geq t_{r'}[\boldsymbol{X}_R](h) \quad \textit{for almost every } h \in \mathcal{H} \tag{2.1}$$

The condition "almost every $h \in \mathcal{H}$" might seem surprising even for someone familiar with the dynamic Wardrop assignment. Recall that the route flow functions $X_r$ are differentiable almost every where and that consequently $x_r$ is defined almost everywhere. In most of the literature, the dynamic Wardrop assignment problem is defined with instantaneous route flows $x_r$ as base variables, rather than with route cumulated flows as base variables. Thus the condition almost everywhere is unnecessary, but the model is less general.

Note that here route travel time models are the only network input, as they completely summarize the dynamic network loading procedure. The following proposition gives a characterization where the network structure appears explicitly.

**Proposition 2.2** (Arc-based characterization of the dynamic Wardrop assignment)**.** *Let:*

- $\boldsymbol{X}_R$ *be an assignment on a dynamic network* $(N, A, (t_a)_{a\in A})$,
- $\boldsymbol{Y}_A = (Y_a)_{a\in A}$ *be the arc cumulated flows resulting from its loading on the dynamic network and* $H_a := \mathrm{id}_{\mathbb{R}} + t_a[Y_a]$.

*Then* $\boldsymbol{X}_R$ *is a dynamic Wardrop assignment if and only if there exists a sequence* $\boldsymbol{H}_{ON} = (H_{on})_{o\in O, n\in N}$ *of continuous increasing functions such that:*

$$H_a \circ H_{on}(h) - H_{om}(h) \geq 0 \quad \forall a = (n, m),\ \forall o, h \tag{2.2}$$

$$\big[H_a \circ H_{on}(h) - H_{om}(h)\big]. \sum_{r\in R_{od}: a\in r} x_r \circ H_{on}(h) = 0 \quad \forall a = (n, m),\ \forall o, h \tag{2.3}$$

The proof of Proposition 2.2 is straightforward. Note that $(H_{on})_{o\in O, n\in N}$ are the earliest arrival functions corresponding to the dynamic network state $\big(N, A, \boldsymbol{H}_A\big)$ as they satisfy Equation (2.2). The dynamic Wardrop principle is embedded in Equation (2.3).

An interesting application of this proposition is to reformulate the dynamic Wardrop assignment problem with arc cumulated flows as primary variables rather than route cumulated flows. The formulation presented here it can be found in (Ran and Boyce, 1996). However, it is a pretty common way to formulate the dynamic Wardrop assignment and a similar formulation can be found in (Leurent, Aguiléra and Mai, 2007). To do so let us first denote $Y_{ad} := \sum_o \sum_{r\in R_{od}} X_r \circ H_{r\prec a}$ the *arc cumulated flows of vehicles originated from o*. Note that $Y_a = \sum_d Y_{ad}$.

**Definition 2.3** (Arc-based formulation of the dynamic Wardrop assignment)**.** *An cumulated flow vector* $\boldsymbol{Y}_A$ *is an arc-based dynamic Wardrop assignment if there exists a sequence* $\boldsymbol{Y}_{AD} = (Y_{ad})_{a\in A, d\in D}$ *of arc cumulated flows and a vector* $\boldsymbol{H}_{ON} = (H_{on})_{o\in O, n\in N}$ *of continuous increasing functions, satisfying the following constraints:*

$$\sum_{d\in D} Y_{ad} = Y_a \qquad \forall a \in A \tag{2.4}$$

$$X_{od} + \sum_{a: a=(n,o)} Y_{ao} = \sum_{a: a=(o,n)} Y_{ad} \qquad \forall od \in OD \tag{2.5}$$

$$\sum_{o\neq n_2} Y_{ad} \circ H_a = \sum_{d, a: a=(n_2,n)} Y_{ad} \qquad \forall a = (n_1, n_2) \tag{2.6}$$

$$H_a \circ H_{on}(h) - H_{om}(h) \;\geq\; 0 \qquad \forall a = (n,m),\ \forall o, h \tag{2.7}$$

$$\left[H_a \circ H_{on_1}(h) - H_{on_2}(h)\right] \,.\, \sum_{d \in D} y_{ad} \circ H_{oa}(h) \;=\; 0 \qquad \forall a = (n_1, n_2),\ \forall o, h \tag{2.8}$$

*letting* $H_a := \mathrm{id}_{\bar{\mathcal{H}}} + t_a[\sum_d Y_{ad}]$.

Definition 2.3 allows to formulate the dynamic Wardrop assignment without making explicit route travel times and route cumulated flows. Equations (2.4), (2.5), and (2.6) are traffic conservation laws. If an arc cumulated flows vector $\boldsymbol{Y}_A$ verifies (2.4), (2.5), and (2.6) then it results from the loading of an assignment of the OD matrix $\boldsymbol{X}_{OD}$. Equation (2.7) defines the earliest arrival time functions and (2.8) states that the arc cumulated flows $Y_{ad}$ are consistent with the shortest routes.

Recall that given a route assignment $\boldsymbol{X}_R$ one can easily construct the corresponding arc cumulated flows decomposed by destination *i.e.* $\boldsymbol{Y}_{AD}$. Conversely from an assignment $\boldsymbol{Y}_{AD}$, one can construct a (non-unique) route assignment $\boldsymbol{X}_R$ by building the tree of possible routes from a given origin to a given destination. The two definitions are equivalent in the following sense:

**Proposition 2.4** (On the equivalence between the formulations)**.** *If $\boldsymbol{Y}_{AD}$ is a solution to the dynamic Wardrop assignment in the sense of Definition 2.3, then a corresponding route assignment $\boldsymbol{X}_R$ is a dynamic Wardrop assignment in the sense of Definition 2.1. The reverse is also true.*

A proof can be found in (Ran and Boyce, 1996, pp. 100-101).

# 2 Variational inequalities

## 2.1 Route-based Formulation

Friesz et al. (1993) were the first to cast a dynamic user equilibrium model as a variational inequality problem. The version we present here is an adaptation of their original formulation for the dynamic Wardop assignment that can be found in (Daniele, Maugeri and Oettli, 1998).

**Proposition 2.5** (Variational formulation of the dynamic Wardop assignment)**.** *Consider a dynamic transport network described by its route travel time models* $\boldsymbol{t}_R = (t_r)_{r\in R}$ *and an OD matrix* $\boldsymbol{X}_{OD}$*. Then* $\boldsymbol{X}^\star_R$ *is Wardrop assignment if and only if it satisfies the following variational inequality:*

$$\sum_r \int_{\mathcal{H}} t_r[\boldsymbol{X}^\star_R](h).\big(x_r(h) - x^*_r(h)\big)dh \geq 0, \quad \textit{for all assignments } \boldsymbol{X}_R \textit{ of } \boldsymbol{X}_{OD} \tag{2.9}$$

*Proof of proposition 2.5 (adapted from Daniele et al.)* **(i)** $\Rightarrow$
Assume $\boldsymbol{X}^*_R$ is a dynamic Wardrop assignment and consider any other assignment $\boldsymbol{X}_R$. Let $\mu_{od}(h) = \min\limits_{r\in R_{o,d}} t_r[\boldsymbol{X}^*_R](h)$. As $\boldsymbol{X}^*_R$ and $\boldsymbol{X}_R$ are both assignments of the same OD matrix:

$$\sum_{r\in R_{od}} x_r(h) = \sum_{r\in R_{od}} x^*_r(h)$$

This yields:

$$\sum_{r\in R_{od}} \int_{\mathcal{H}} (t_r[X^*](h) - \mu_{od}(h)).(x_r(h) - x^*_r(h))dh$$

$$= \sum_{r\in R_{od}} \int_{\mathcal{H}} (t_r[X^*](h)).(x_r(h) - x^*_r(h))dh$$

It is enough to show that the right hand side of the previous equation is positive. Now $\big(t_r(X^*(h)) - \mu_{od}(h)\big).\big(x_r(h) - x^*_r(h)\big) \geq 0$ by definition of dynamic Wardrop assignment. Thus the result.

**(ii)** $\Leftarrow$

Let $\boldsymbol{X}^*_R$ be an assignment satisfying (2.9). Define $\mu_{od}(h)$ as previously. The proof proceeds by contradiction. Assume that the Wardrop principle is not satisfied for a route $r_1 \in R_{od}$. Then the set $S = \big\{h \in \mathcal{H} : x^*_{r_1}(h) > 0 \text{ and } t_{r_1}(\boldsymbol{X}^*)(h) > \mu_{od}(h)\big\}$ is of non null measure.

Let us now apply the inequality to an assignment $\boldsymbol{X}_R$ defined as follows. $\boldsymbol{X}_R$ is differing rom $\boldsymbol{X}^*_R$ only on $S$ and on the routes $r \in R_{od}$. Moreover $x_{r_1}(h) = 0$ over $S$ and $x_r(h) = x^*_r(h)$ if $(r, h)$ is such that $t_r[\boldsymbol{X}^*](h) \neq \mu_{od}(h)$. One easily verifies that such an $\boldsymbol{X}_R$ exists. When replacing in (2.9):

$$\begin{aligned}
&\sum_r \int_{\mathcal{H}} t_r[X^*(h)].\big(x_r(h) - x_r^*(h)\big)dh \\
&= \sum_{r\in R_{od}} \int_{\mathcal{H}} \big(t_r[X^*(h)] - \mu_{od}(h)\big).\big(x_r(h) - x_r^*(h)\big)dh \\
&= \int_{\mathcal{H}} \big(t_{r_1}[X^*(h)] - \mu_{od}(h)\big).\big(x_{r_1}(h) - x_{r_1}^*(h)\big)dh \\
&= \int_{\mathcal{H}} \big(t_{r_1}[X^*(h)] - \mu_{od}(h)\big).\big(-x_{r_1}^*(h)\big)dh < 0
\end{aligned}$$

Thus the result. □

## 2.2 Arc-based Formulation

The following proposition is due to Chen and Hsueh (1998).

**Proposition 2.6** (Arc-based variational formulation)**.** *An arc cumulated flow vector $\boldsymbol{Y}_A^\star$ is an arc-based Wardrop assignment if and only if it satisfies the quasi-variational inequality:*

$$\int_{\mathcal{H}} \sum_{d,a=(n_1,n_2)} \sum t_a[Y_a^\star].(y_a - y_a^\star) \geq 0 \quad \forall\, \boldsymbol{Y}_A \in \Omega(\boldsymbol{Y}_A^\star) \tag{2.10}$$

*where $\Omega(\boldsymbol{Y}_A^\star)$ is the set of $\boldsymbol{Y}_A$ such that there exists $\boldsymbol{Y}_{AD}$ satisfying the following constraints:*

$$\sum_{d\in D} Y_{ad} \;=\; Y_a \qquad \forall a \tag{2.11}$$

$$X_{md} \;+ \sum_{a:a=(n,m)} Y_{ad} \circ (\mathrm{id}_{\bar{\mathcal{H}}} + t_a[Y_a^\star]) \;=\; \sum_{a:a=(m,n)} Y_{ad} \qquad \forall d \in D,\ m \in N : m \neq d \tag{2.12}$$

The proof of the proposition consists in showing that the solutions of Proposition 2.6 are solutions to the (dynamic) Wardrop assignment problem as exposed in Definition 2.3. A well-written proof can be found in (Bliemer and Bovy, 2003).

The following points are noteworthy:

- Inequation (2.10) is not a variational inequality as such but a *quasi-variational inequality* since the set of auxiliary variables $\Omega$ depends on $\boldsymbol{Y}_A^\star$. This has consequences in the algorithmic design.

- Arc-based formulations are especially interesting in large networks where enumerating all the possible routes serving a pair $OD$ is a tedious tasks.

## 2.3 Algorithms

Here we will focus on the algorithms that adress the arc-based variational inequality. Route-based algorithms are essentially variants of the route swapping algorithm that is presented in the next section.

**Philosophy.** Most of the algorithms proposed to solve the quasi-variational inequality of Proposition 2.6 rely on a *nested relaxation method* or *nested projection method*. They decompose the problem in two conceptual steps: solving the variational inequality for a fixed the set of auxiliary variables $\Omega$ and then dealing with the general problem where $\Omega$ depends on the candidate solution $\boldsymbol{Y}_A^\star$.

The *relaxation method* (also known as diagonalisation method) is a standard technique to solve variational inequality problems (for a review see Patriksson, 1999). In a nutshell, the relaxation method cast a variational inequality into a sequence of subproblems which are, in general, non-linear programming problems.

To deal with the *quasi*-variational nature of the inequality, it is necessary to solve a sequence of regular variational inequalities, which is why those methods are said to be *nested*. At each iteration the solution changes, thus inducing a new set of auxiliary variables and yielding the new variational inequality to solve.

To sum up a two loop algorithm has to be designed. At each step in the outer loop, the current solution $\boldsymbol{Y}_A^{n,\star}$ is updated and so is the set $\Omega := \Omega(\boldsymbol{Y}_A^{n,\star})$. In the inner loop, the variational inequality obtained by considering (2.10) on $\Omega$ and not on $\Omega(\boldsymbol{Y}^\star)$ is solved. This can be achieved by an iterative procedure inspired by the relaxation method, where $\boldsymbol{Y}^{n,\star}$ is progressively approximated by a sequence of cumulated flows $\boldsymbol{Y}^k$. Other possibilities include projection methods.

**Details of an algorithm.** We are going to present a nested relaxation method due to Chen and Hsueh (1998). To state it, it is first necessary to discretize the quasi-variational inequality. We reinterpret the previous notations as follows. The set of departure times is now a finite set of integers, $\mathcal{H} = [1; H]$ and $y_a$ are $H$-dimensional vectors, denoted $y_a = (y_a^1, \ldots, y_a^H)$. We restate arc travel time models as functions of the (discretized) instantaneous flows $y_a$ rather than the cumulated flows, and they are assumed to be integer-valued. In the same order of idea, travel time models are denoted $\boldsymbol{y}_A := (y_a^1, \ldots, y_a^H) \mapsto t_a[y_a^1, \ldots, y_a^H] = \left(t_a^h[y_a^1, \ldots, y_a^H]\right)_{h\in\mathcal{H}}$ and the travel time vector is now $\boldsymbol{t}_A := (t_a^h[\boldsymbol{y}_A])_{a\in A, h\in h}$. Assuming that the causality principle is respected by the arc travel time models, we can write:

$$t_a[y_a^1, \ldots, y_a^H](h) = t_a[y_a^1, \ldots, y_a^h](h) \quad \forall a$$

The problem stated in Proposition 2.6 can then be straightforwardly discretized by replacing the time integration signs by sum signs. This yields:

$$< \boldsymbol{t}_A[\boldsymbol{y}_A^\star], (\boldsymbol{y}_A - \boldsymbol{y}_A^\star) >\geq 0 \quad \forall\ \boldsymbol{y}_A \in \Omega(\boldsymbol{y}_A^\star) \tag{2.13}$$

where $< .,. >$ denotes the standard scalar product in a real vector space and $\Omega(\boldsymbol{y}_A^\star)$ is the set of flows defined by Equations (2.11-2.12), adapted for a discrete-time setting and restated in terms of instantaneous flows.

Recall that to deal with the quasi-variational nature of inequality (2.10), we first fix the set of auxiliary variables to $\Omega(\boldsymbol{y}_A^{\star,n})$, where $\boldsymbol{y}_A^{\star,n}$ is a flow vector that is updated at each iteration of the outer loop. The relaxation procedure then consists in relaxing most of the dependencies of the travel time vector More precisely, for each coordinate $t_a^h\big[\boldsymbol{y}_A^\star\big]$ of the travel time vector $\boldsymbol{t}_A\big[\boldsymbol{y}_A^\star\big]$ we are going to fix all the arc flows to the value of $\boldsymbol{y}_A^k$ except for $y_a^h$. Thus in the relaxed version of the variational inequality, the travel time models are replaced by the expression:

$$t_a^h\left[\boldsymbol{y}_A^k, \boldsymbol{y}_A\right] = t_a^h\big[\underbrace{y_a^{k,1}, \ldots, y_a^{k,h-1}}_{\text{current state}}, \quad \underbrace{y_a^h}_{\text{new flows}} \quad \big]$$

which yields the following variational inequality:

$$< \boldsymbol{t}_A[\boldsymbol{y}_A^k, \boldsymbol{y}_A^\star], (\boldsymbol{y}_A - \boldsymbol{y}_A^\star) >\geq 0 \quad \forall\ \boldsymbol{y}_A \in \Omega(\boldsymbol{y}_A^{\star,n})$$

Now this variational inequality problem can be shown to be equivalent to:

$$\min_{\boldsymbol{y}_A\in\Omega(\boldsymbol{y}^{\star,n})} Z(\boldsymbol{y}_A^k, \boldsymbol{y}_A) = \sum_h \sum_a \int_0^{y_a^h} t_a^h\big[y_a^{k,1}, \ldots, y_a^{k,h-1}, x\big]\, dx \tag{2.14}$$

The program (2.14) is a convex optimization program as soon as the maps $x \mapsto t_a^h\left[y_a^{k,1}, \ldots, y_a^{k,h-1}, x\right]$ are strictly increasing. It can be solved by classic non-linear programming techniques such as the Frank-Wolf algorithm, which is well-known and widely used in the transport science community.

The overall algorithm is presented in pseudo-code in Algorithm 2.1.

**Algorithm 2.1** `NestedRelaxationMethod`$(\boldsymbol{t}_A, \boldsymbol{x}_{OD}, \mathcal{H})$

**Inputs:** $\mathcal{H} = [1; H]$, the set of departure times
$\boldsymbol{x}_{OD} = (x_{od}^h)_{od \in OD, h \in \mathcal{H}}$, an OD matrix of time-varying flows
$\boldsymbol{t}_A = (t_a^h)_{a \in A, h \in \mathcal{H}}$, a vector of arc travel time models
**Outputs:** An arc instantaneous flow vector $\boldsymbol{y}_A = (y_a^h)_{a \in A, h \in \mathcal{H}}$
**Parameter:** $w_n$ a decreasing sequence from 1 to 0.
**Initialization.** $n := 0$. Set $\boldsymbol{y}_a^k$ with any heuristic assignment procedure such as all or nothing assignment or incremental assignment. Set $\tau_a^h := t_a^h[0, \ldots, 0]$ for all $a$ and $h$.
**Outer loop: Do**
**Set** $n := n + 1$ and $\boldsymbol{y}_a^{\star,n} := \boldsymbol{y}_a^k$
**Update** $\tau_a^h := (1 - w_n)\tau_a^h + w_n t_a^h[\boldsymbol{y}_a^{\star,n}]$ for all $h$ and $a$
**Set** $k := 0$, $\boldsymbol{y}_a^k := \boldsymbol{y}_a^{\star,n}$ and $\Omega := \Omega(\boldsymbol{y}_a^{\star,n})$
**Inner loop: Do**
**Solve** the optimization program (2.14) on $\Omega$ by any suitable method.
(*e.g.* the FW method)
**Set** the results to $\boldsymbol{y}_a^{k+1}$.
**Until** $\boldsymbol{y}_a^{k+1} \approx \boldsymbol{y}_a^k$
**Until** $\tau_a^h \approx t_a^h[\boldsymbol{y}_a^{\star,n}]$ for all $h$ and $a$

**Variants and Convergence.** Variants of the latter relaxation method for the resolution of the arc-based quasi-inequality have been proposed by Ran and Boyce (1996). Projection methods have been proposed by Bliemer and Bovy (2003) and Szeto and Lo (Lo and Szeto, 2002; Szeto and Lo, 2004).

This category of algorithms has been tested extensively on small networks (less than one hundred arcs) and the algorithms showed reasonable convergence. However, to our knowledge there has been no large size implementation of such methods. This is quite surprising since the rationale behind arc-based formulation is to avoid route enumeration, which is typically infeasible for large networks.

# 3 Fixed point problems

## 3.1 A route-based formulation and its applications

**Statement.** The formulation exposed here is based on route swapping processes. A route swap process can be informally interpreted as a re-routing strategy of users in a non-equilibrium state given a route travel time pattern. For instance assigning all the users in an all or nothing fashion is a form of route swapping process, although very crude. Obviously there is a wide collection of possible route swapping processes (for a review see Mounce and Carey, 2010), and although the ones exposed below are fairly realistic they have no empirical validity. Here we consider the route swapping as a theoretical and algorithmic device so this question is out of our scope.

Let us now precisely define some commonly used route swapping processes. Formally a route swapping process is a function from the set of the possible assignments of an OD matrix $X_{OD}$ into itself. The following notations will be useful. For $r, r' \in R$, $\boldsymbol{\delta}_{rr'}$ is the vector that has $-1$ in the $r$-th coordinate, 1 in the $r'$-th coordinate and 0 elsewhere. $(.)_+$ is the positive part and $r \sim r'$ means that the routes $r$ and $r'$ connect the same OD pair.

**Definition 2.7** (weighted pairwise swapping)**.** *A weighted pairwise swapping process is a function $\boldsymbol{X}_R \mapsto RS[\boldsymbol{X}_R]$ such that:*

$$RS[\boldsymbol{X}_R](h) = \boldsymbol{X}_R(h) + \alpha \sum_{r,r':r\sim r'} X_r(h)\Big(t_r[\boldsymbol{X}_R](h) - t_{r'}[\boldsymbol{X}_R](h)\Big)_+ \boldsymbol{\delta}_{rr'}$$

*where $\alpha$ is a (small) positive real.*

In weighted pairwise swapping, flow transfers occur between each pair of route connecting the same OD where one of the routes is longer. The swap rate is proportional to the flow on the longer route multiplied by the travel time difference on the two routes. Pairwise swapping has been introduced by Smith and Wisten (1995). Note that $\alpha$ needs to be chosen sufficiently small to ensure that $RS[\boldsymbol{X}_R](h)$ stays positive.

**Definition 2.8** (weighted shortest route swapping)**.** *The weighted shortest route swapping process is the function $\boldsymbol{X}_R \mapsto RS[\boldsymbol{X}_R]$ such that:*

$$RS[\boldsymbol{X}_R](h) = \boldsymbol{X}_R(h) + \alpha \sum_{r} \sum_{r' \in R^r_{min}(h)} \frac{X_r(h)\Big(t[X_r](h) - t[X_{r'}](h)\Big)_+}{|R^r_{min}(h)|} \boldsymbol{\delta}_{rr'}$$

*where $R^r_{min}(h)$ is the set of early arrival paths when departing at $h$ on the OD pair that connects $r$ and $\alpha$ is a (small) positive real.*

In weighted shortest route swapping, flow transfers are still proportional to the flow on the longest route multiplied by the travel time difference but here swapping occurs only towards the shortest paths on each OD pairs.

**Definition 2.9** (unweighted shortest route swapping)**.** *The unweighted shortest route swapping process is the function $\boldsymbol{X}_R \mapsto RS[\boldsymbol{X}_R]$ such that:*

$$RS[\boldsymbol{X}_R](h) = \boldsymbol{X}_R(h) + \alpha \sum_r \sum_{r' \in R^r_{min}(h)} \frac{X_r(h)}{|R^r_{min}(h)|} \boldsymbol{\delta_{rr'}}$$

*where $R^r_{min}(h)$ is the set of shortest paths when departing at $h$ on the OD pair that connects $r$ and $\alpha$ is a positive real between 0 and 1.*

In unweighted shortest route swapping, flow transfers are not proportional to the difference in travel times anymore and occur only towards the shortest paths on each OD pairs.

An important (and straightforward) property of the route swapping processes we presented is that their fixed points are the Wardrop equilibriums of the dynamic network represented by $\boldsymbol{t}_R$ and reversely. This can easily be seen by noting that the second terms in the definition of each swapping process is some kind of measure of the users' incentives to change route.

**An existence result.** Using this formulation, an existence result can be established using the Schauder fixed point theorem which is a generalization of the one of Brouwer for infinite dimension spaces. Such a proof is presented by Mounce (2003) assuming given route travel models that are continuous for a certain topology on the set of flows. It is then shown in (Mounce, 2007) that the route travel time models arising from networks with bottleneck travel time models are indeed continuous.

**The route swapping algorithm: statement and convergence results.** A natural algorithm for finding the fixed point of a function is to iteratively apply this function until convergence. Obviously this convergence is only guaranteed under certain restrictive assumptions. Up to now no such result exist and the route swapping algorithm has to be viewed as heuristic only.

Yet, route swapping algorithms are widely used methods especially in micro-simulation based dynamic network models (see for instance the work of Mahmassani and colleagues: Mahmassani et al., 1995; Jayakrishnan, Mahmassani and Hu, 1994; Hu and Mahmassani, 1997). Among their empirical findings is that $\alpha$, the route swap parameter, as well as the route swapping process greatly influences the quality of the results. In particular the unweighted shortest route swapping is inefficient compared to the others and the resulting algorithms tends to have oscillating behaviours. Large scale implementations have been proposed and shows good convergence results, although this last point has to be mitigated due the poor convergence measures used in these studies.

## 3.2 Arc-based formulations and related algorithms

**Statement.** The formulation presented here is due to Leurent (2003*b*) in the context of the LADTA model. Let us introduce the following notations for a given assignment problem characterized by a dynamic network $G = (N, A, \boldsymbol{t}_A)$ and an OD matrix $\boldsymbol{X}_{OD}$.

- The *loading function* associates with a route cumulated flow vector the arc flows resulting from the resolution of the corresponding dynamic network loading problem. It is denoted $\boldsymbol{X}_R \mapsto F_S[\boldsymbol{X}_R] = \boldsymbol{X}_A$.

- The *constrained loading function* that associates with a route flow vector and an arc travel time function vector, the arc flows obtained by simply translating the route cumulated flows by the corresponding route travel time functions. It is denoted $\boldsymbol{X}_R \mapsto \tilde{F}_S[\boldsymbol{X}_R; \boldsymbol{\tau}_A] = \boldsymbol{X}_A$.

- The *shortest route function* that associates to each travel time function vector the corresponding earliest arrival time functions. It is denoted $\boldsymbol{\tau}_A \mapsto F_{SR}[\boldsymbol{\tau}_A] = \boldsymbol{H}_{ON}$.

- The *user function* that gives the set of possible assignments $\boldsymbol{X}_R$ of $\boldsymbol{X}_{OD}$ on the shortest routes. $\boldsymbol{H}_{ON} \mapsto F_D[\boldsymbol{H}_{ON}] = \{\boldsymbol{X}_R\}$.

Note that Equation (2.8) in Definition 2.3 rewrites $\boldsymbol{X}_R \in F_D[\boldsymbol{H}_{ON}]$, that the vector $\boldsymbol{H}_{ON}$ satisfying Equation (2.7) is exactly $F_{SR}[\boldsymbol{t}_A[\boldsymbol{Y}_A]]$ and the arc flow vector such that there exists a flow vector $\boldsymbol{Y}_{AR}$ is unique and is exactly

$\boldsymbol{X}_R \mapsto F_S[\boldsymbol{X}_R] = \boldsymbol{Y}_A$. Thus the dynamic Wardrop assignment problem is equivalent to the following fixed point problem:

$$\textit{Find } \boldsymbol{Y}_A \textit{ such that: } \boldsymbol{Y}_A \in F_S \circ F_D \circ F_{SR} \circ \boldsymbol{t}_A[\boldsymbol{Y}_A] \tag{2.15}$$

Now by definition of the loading procedure $F_S[\boldsymbol{X}_R] = \tilde{F}_S\Big[\boldsymbol{X}_R; \boldsymbol{t}_A \circ F_S[\boldsymbol{X}_R]\Big]$, so we have the following formulation:

**Definition 2.10** (Arc-based fixed point formulation)**.** *Find* $\boldsymbol{Y}_A$ *such that:*

$$\boldsymbol{Y}_A \in \tilde{F}_S\Big[F_D \circ F_{SR} \circ \boldsymbol{t}_A[\boldsymbol{Y}_A]; \boldsymbol{t}_A[\boldsymbol{Y}_A]\Big] \tag{2.16}$$

It is important to understand the difference between the formulation in Equation (2.15) and Definition 2.10. Essentially, in the first formulation in addition to the demand-supply circular dependency, there is a second circular dependency between arc flows and route travel times for given route flows, accounting for the dynamic network loading problem. On the contrary, in the latter formulation, the dynamic Wardrop assignment is expressed as a single fixed point problem. The dependencies between the main variables of the problems are depicted in Figure 2.1.

## 3.3 A convex combination algorithm.

The algorithm proposed by Leurent (2003*b*) to solve the arc-based fixed point formulation is a classic convex combination algorithm. It is presented in pseudo-code in Algorithm 2.2. When the parameter $w_k = 1/k$ the method is termed the *method of successive averages.*

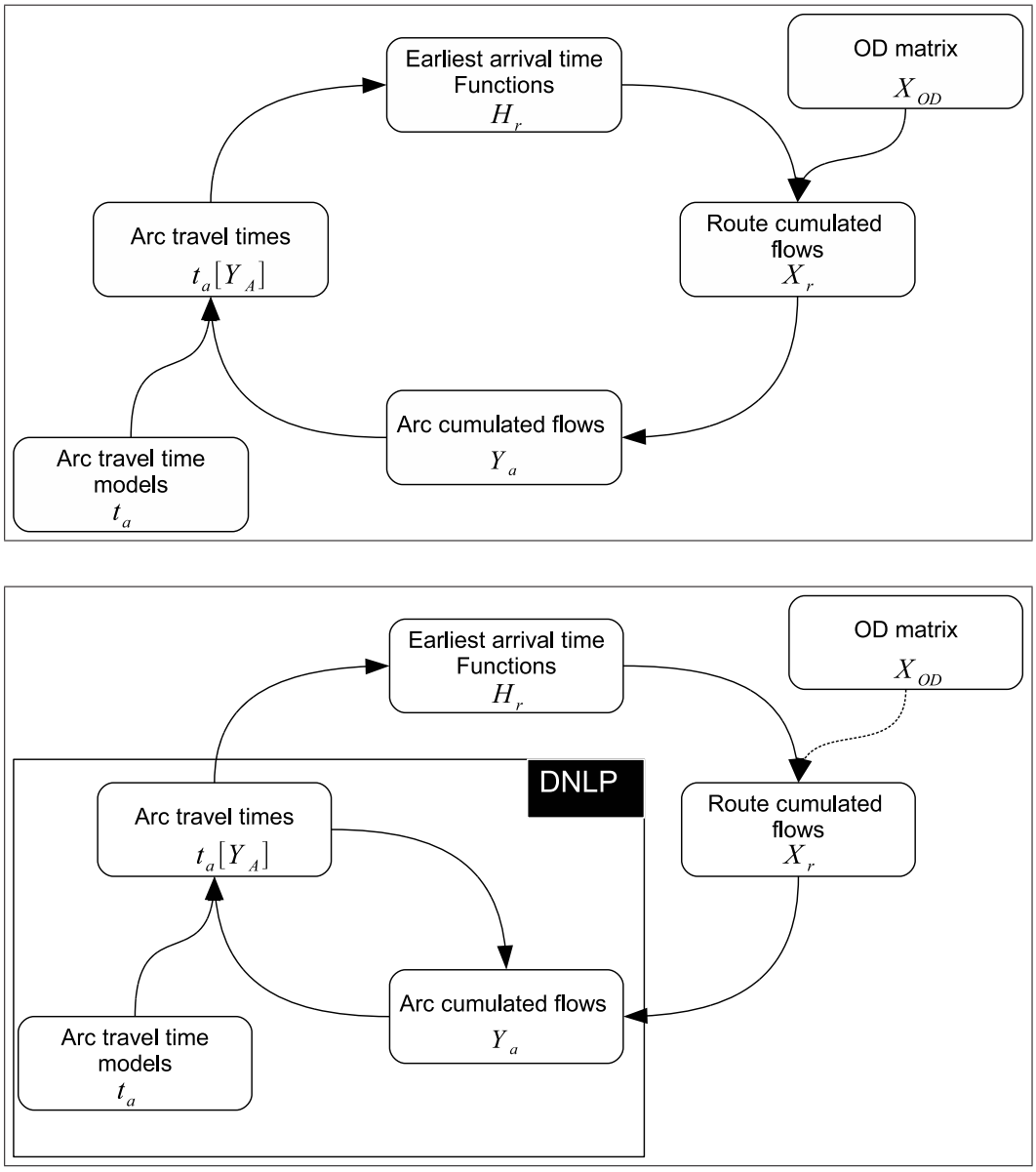



Figure 2.1: Relationship structure in the dynamic Wardrop assignment (Figure of the top adapted from (Leurent, 2003*b*))

**Algorithm 2.2** `LADTA RouteChoice`$(\boldsymbol{X}_C)$

**Inputs:** An $OD$ matrix $\boldsymbol{X}_{OD}$
**Outputs:** $\boldsymbol{Y}_A$ the arc cumulated flows for each user category.
**Parameter:** $w_k$ a decreasing sequence from 1 to 0.
**Initialize** $\boldsymbol{Y}_A^{[0]} := 0$ and $k := 0$
**Repeat**
$\boldsymbol{\tau}_A := t_A[\boldsymbol{Y}_A^{[k-1]}]$
$\boldsymbol{H}_{ON} := F_{SP}(\boldsymbol{\tau}_A)$
$\boldsymbol{Y}_A := F_D(H_{ON}; \boldsymbol{X}_{OD})$
$\boldsymbol{Z}_A := \tilde{F}_L(\boldsymbol{X}_R, \boldsymbol{\tau}_A)$
$\boldsymbol{Y}_A^{[k]} := w_k.\boldsymbol{Y}_A^{[k-1]} + (1 - w_k).\boldsymbol{Z}_A$
**Until** $\boldsymbol{Y}_A^{[k]}$ satisfies a certain criterion.
**End For**

Algorithm 2.2 has been tested in details on small networks (Leurent, Mai and Aguilèra, 2006), as well as numerous variants of the previous computation scheme. For the occasion they have developed advanced convergence criteria in order to provide a detail assessment of the convergence property of the algorithm. The results look promising. Numerical experiences on very large networks (approx. 10.000 arcs in Aguiléra and Leurent, 2009) have shown that the algorithm could provide a reasonable solution to the dynamic Wardrop assignment problem in a reasonable computation time.

Bellei et al. (2005) developed a similar algorithm, although for a slightly different model where a stochastic user equilibrium paradigm is used. The algorithm has been tested on both route-based and arc-based fixed point formulations. For the first formulation the dynamic network loading problem is solved by applying the convex combination procedure presented in subsection 3.4 of the previous chapter. Numerical experiments have shown that the second formulation achieved similar degrees of convergence in many time less than when using the first formulation. This result is not surprising as the first formulation is a bi-level problem. A more surprising fact is that the number of iterations required for both algorithms is approximately the same.

A method of successive averages was proposed by Tong and Wong (2000) for a route-based formulation where the loading problem is solved with an heuristic based on a decomposition of the traffic flows in "platoons" of vehicles.

# 4 Extensions to dynamic user-equilibrium problems

In the simple dynamic Wardrop assignment, only one dimension of travel is taken in account: which routes to undertake. The range of travel decisions is in fact much wider and includes whether or not to travel, at which time to start the trip and possibly intermediate stops on the way to perform activities. Taking account of those decision leads to consider more general *Dynamic User Equilibrium* models (DUE models).

When it comes to the introduction of advanced choice model in dynamic transport modelling two trends can be observed. *Sequential approaches* are essentially a variant of four-stage models and where a simple dynamic Wardrop assignment is used instead of a static one. The choice models and the assignment procedures are sequentially applied with feedback until some sort of convergence if any. On the contrary *integrated models* formalize the transport supply and demand model in a unique framework and clearly specify a user equilibrium principle. In this thesis we focus on the latter.

## 4.1 User equilibriums with generalized costs and multi-class users modelling

**Trade-off between time and money.** A first generalization of the dynamic Wardrop equilibrium is to take account of the monetary costs incurred by the users. This requires to model the possible trade-offs of users between time and money. This is generally achieved using the economic theory of consumers. In this framework, consumers (in our case the transport network users) are assumed to maximize their utilities subject to a budget constraint. The introduction of time-money trade-offs was discussed thoroughly in the economic literature starting by the paper of Becker (1965). Becker considered time as a necessary input to consume goods but may also be assigned to work hours which results in an increase of income by the mean of a fixed hourly wage. The budget of time within a day being constrained to 24 hours, the optimum for the users is obtained when time is valued at the fixed hourly wage. More complex value of time models can be established leading to more subtle definition of the value of time (DeSerpa, 1971; Evans, 1972), that notably varies with the type of activities undertaken.

These theoretical results suggested the introduction of a *generalized cost*

*of travel* including both monetary costs and travel time costs, the latter being expressed in monetary terms thanks to a value of time. Denoting $t$ and $p$ the travel time and monetary costs of a trip, the generalized cost incurred by a user with value of time $\nu$ is thus:

$$g(t, p; \nu) = \nu.t + p$$

As stated earlier $\nu$ might depend on the trip purpose and on the user's income.

**Representing user classes.** We have already seen that users can be segmented according to the traffic performance of their vehicles and their driving behaviour. The previous paragraph also shows that they might differs according their values of time and the next section introduces preferred arrival times that might vary from within the users populations.

This leads to the concept of *user classes*. A user class is a set of characteristics summarizing all the relevant data about a type of users. For instance Bliemer (2001) suggests that user should be categorized according their vehicle types (*e.g.* passenger cars, trucks and vans...), the driver characteristics (ability to drive, economic preferences such as his value of time...), their network access restriction (to take in account dedicated road infrastructure such truck or high-occupancy vehicle lanes), purpose of travel and level of information.

From a formal perspective the segmentation can be addressed in two manners. *Discrete sets* of user classes might be considered with a population associated to each of these. A second option is to introducing *continuous user classes* by allowing some characteristics to take all the possible values within a real interval. The user population repartition among the classes is then described by the mean of distributions over the set of the characteristics. For instance one might want to describe a user population where the value of time is distributed according a log-normal law.

**Formulation and algorithms.** With discrete users classes the formulations and algorithmic approaches presented earlier can be straightforwardly extended. The new user equilibrium condition is the same as the one of Definition 2.1, replacing route travel times by route generalized costs. Importantly the variational inequality formulation can be extended to deal with

the new equilibrium. However, the resulting algorithms tends to be less stable and have slower convergence.

*Continuous user classes* have been commonly used in static models but are nearly absent from dynamic models. To our knowledge the only exceptions are the analytical models dealing with DUE with departure time choice on small one- or two-arcs networks. They are presented in the following subsection.

## 4.2 Departure time choice modelling

**Vickrey's model.** In Vickrey's model a set of commuters wish to reach a central business district accessible by one route with bounded capacity. Each user is characterized by a preferred arrival instant $h_p$ and assesses the decision of departing at an instant $h$ using a cost function of the form:

$$G(h;t(h)) = \underbrace{\nu.t(h)}_{\text{traversal costs}} + \overbrace{\alpha.(h+t(h)-h_p)_- \ + \ \beta.(h+t(h)-h_p)_+}^{\text{schedule delay costs}} \tag{2.17}$$

where:

- $t(h)$ is the travel time on the route when departing at $h$,
- $\nu$ is the value of time of the commuter,
- $\alpha$ [resp. $\beta$] is the marginal cost of arriving earlier [resp. later] than preferred,
- $(.)_+$ and $(.)_-$ stand for the positive and negative part of the scedule delay.

There are two standard ways of describing the set of users. Either one considers a finite number of categories of commuters, differing by their preferred arrival times (and also possibly their value of time, value of arriving late or early); or one considers that users have preferred arrival instants distributed among a set of possible values. In the later case, the"S-shape" assumption is made: there is a single interval during which the density of commuters exceeds capacity. This assumption makes the model analytically tractable, and induces a travel time pattern similar to the one with a unique $h_p$ shared by all commuters.

When $t$ is assumed to follow a bottleneck travel time model, an equilibrium departure time distribution can be found. The typical equilibrium situation is depicted in Figure 2.2 for a bottleneck of capacity $k$. The set of users is modelled by a cumulative distribution $X_p$ over the set of their preferred arrival instants. Their choices of departure instants are given by the cumulative distribution $X_+$. The bottleneck model allows one to compute the cumulative distribution $X_-$ of users at the exit of the bottleneck.

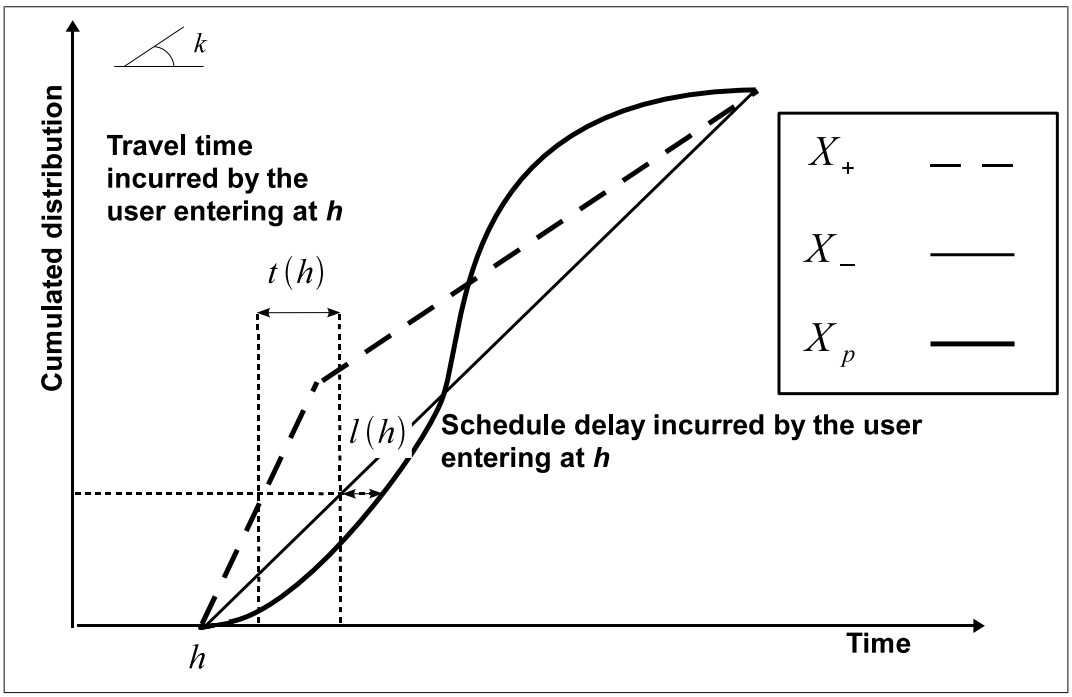



Figure 2.2: Vickrey's bottleneck model

Figure 2.2 can be interpreted as such: the horizontal difference between $X_+$ and $X_-$ (*i.e.* $t(h)$) gives the amount of time needed to traverse the bottleneck, when entering the bottleneck at instant $h$. The horizontal difference between $X_-$ and $X_p$ (*i.e.* $l(h)$) gives the schedule delay at arrival. The travel time function $t$ is a piecewise function with only two admissible slopes and a single maximum. It increases at the beginning of the congestion period, when users are arriving earlier than preferred. When $t$ is decreasing, users are arriving later than preferred. Note that the simple form of the schedule delay cost function (*i.e.* the two last terms of Equation 2.17) implies the piecewise

linear shape of the travel time function. Under more general assumptions, it would be smoother.

Since Vickrey, the transportation community has investigated the field in two main directions. Some works, mainly from transport economists, have focused on users heterogeneity. Others have proposed extensions to whole networks.

**Trip scheduling with users heterogeneity** Heterogeneity in preferred arrival times can be addressed either in a discrete manner, by allowing only a finite set of preferred arrival times, or continuously using a distribution. The finite case has been studied extensively by Lindsey (2004) while the continuous case was first treated by Hendrickson and Kocur (1981). Heterogeneity pertaining to the costs of travel time and of schedule delay has been studied, among others, by Arnott, de Palma and Lindsey (1993) and Van der Zijpp and Koolstra (2002). Other extensions include the modelling of stochastic demand and capacity, multiple routes or elastic demand (see Arnott et al., 1998, for a review). When users are at equilibrium, the bottleneck model predicts a congestion pattern with a single peak in travel time. In the numerous extensions, the resulting congestion patterns are very similar to the homogeneous case. When considering a finite number of preferred arrival instants, there is a limited number of peaks in travel time (at most one per preferred arrival instant) and a spontaneous segregation among users is observed. Commuters with different preferred arrival instants depart at different instants (Lindsey, 2004). The case where users' preferences are distributed over an interval has received less attention. Papers in this line mainly considered "S-shape" distribution (Smith, 1984; Daganzo, 1985). As exposed in the previous paragraph this case is practically equivalent to the one with a single preferred arrival time and produces exactly the same travel time pattern.

**DUE with departure time choice on networks** Vickrey's model has been extended to networks, in an attempt to produce operational planning models. The computation of the user equilibrium in such a context is known as the DUE problem with departure time choice. Friesz et al. (1993) first proposed a formulation of the user equilibrium with both route and departure time choice as a variational inequality. Their model considers users dispatched among several origin-destination pairs, with a unique preferred

arrival time by OD. Since then most of the models proposed in the literature rely on Friesz's original paradigm (*e.g.* Wie, Tobin and Carey, 2002). Rather than focusing on users heterogeneity, this part of the literature has made considerable efforts to improve congestion representation by integrating sophisticated traffic models.

To our knowledge the only network model considering distributions of preferred arrival times is by Bellei et al. (2005). Their approach is stochastic and uses an extension of Vickrey's model given by the following stochastic continuous logit model (see Ben-Akiva and Lerman, 1985)). The probability for a user to choose to arrive in the interval $[h, h + \Delta h]$ is given by:

$$P(h; t_r)\Delta h = \frac{\exp\big(-g(h; t_r)/\mu\big)}{\displaystyle\int_{\mathcal{H}} \exp\big(-g(h; t_r)/\mu\big)\, du}\Delta h \tag{2.18}$$

where $\mu$ is the *heterogeneity parameter*. A similar approach is used in the METROPOLIS model developed by de Palma and colleagues (*e.g.* De Palma and Marchal, 2002) but applied to a model with a single preferred arrival time window shared by all users. The rationale for such a continuous logit model is essentially to ease computation processes and is not justified by behavioural considerations. Indeed in transport science the use of logit model is more frequent for *discrete* choice modelling where a stochastic approaches allows to avoid "step-like" behaviours.

# Conclusion

In this chapter, different formulations and algorithms for the dynamic traffic assignment problem were reviewed. It was shown that dynamic traffic assignment models can be stated in a rather unified framework and that in this framework some mathematical results already exist. Although there are still some work to do to establish a solution method with theoretical guarantees regarding convergence, empirically efficient algorithms exists.

However, when it comes to more complex dynamic user equilibrium models, there is a clear lack of a common framework.

# Part II

# Dynamic congestion games: a general model and its application to dynamic traffic assignment

# Part introduction

The bibliographic review presented in Part I showed that the standard model of dynamic traffic assignment is a mature topic both regarding its formulation and the solution methods developed to solve it. It is not the case of more complex dynamic user equilibrium models that incorporate refined representations of the transport demand.

A possible explanation is that existing formulations of the dynamic user equilibrium often quote Nash equilibrium as the equilibrium concept but never formally express it as a game. This is quite surprising as the behavioural assumptions retains for the great majority of DUE models, *e.g.* users have perfect information and are selfish cost minimizing agents, are also the one of the Nash equilibrium. The question of whether current DUE models can be formalized as Nash games is thus fundamental: if it can not, then it is important to understand why and what it means from a behavioural perspective; if it can, then the numerous results from game theory may be applied.

## Objectives

1. *To set up a general framework for analytical DUE models.* By general it is meant here that no specific travel time models will be used and that models of demand including departure time choice and multi-class user models will be allowed. The framework proposed is not compatible as such with an activity-based approach but we will discuss that it provides a suitable starting point to provide a true framework for

2. *To provide an existence result for a user equilibrium in this model.* The

corresponding theorem is an efficient tool to show existence in current models due to the generality of the framework.

3. *To formulate the dynamic user equilibrium as a Nash equilibrium.*

## Structure

This part is organized as follows. Chapter 3 presents the so-called dynamic congestion games are presented and an existence result for dynamic congestion games is proved. Then, in Chapter 4, it is shown that the dynamic Wardrop assignment problem can be formulated as a dynamic congestion game and that the existence theorem allows to show previously established existence results.

Chapter 3

# Dynamic congestion games: presentation and a simple illustration

Consider over a time interval, say a day, a network prone to congestion. A set of users travel along directed paths, called routes, connecting origins to destinations. At the beginning of the day users are at origins and wish to reach a specific destination by the end of the day. In order to do so, they make a travel decision on the network, *i*.e. choose a route and a departure time. Yet users' decisions depend on route travel time over the network, itself depending on the flow of users taking each route and thus on the decisions of the other users.

Finding an equilibrium (in the Nash sense) of such a problem is, roughly speaking, the Dynamic User Equilibrium (DUE) problem. Up to now few theoretical results have been established regarding the DUE problem (with the notable exceptions of Mounce (2006; 2007) and Zhu and Marcotte (2000) all in the area of dynamic Wardrop assignment), and no general existence result is known.

The purpose of this chapter is to propose a suitable framework in which to study this problem and to give a general equilibrium result that covers most of the previous ones.

We model our problem as a game. In our search for a framework for the dynamic user equilibrium we naturally define a new class of games: *dynamic congestion games*.

## Dynamic congestion games

Let $G = (N, A)$ be a directed graph. We define a *dynamic congestion game* to be a nonatomic (in the sense of Mas-Colell, 1984) game in which each user chooses a *route* (an acyclic directed path), together with a departure time. Denoting by $R$ the set of routes and by $\mathcal{H}$ the set of admissible departure times, the user's possible strategies are in $\mathcal{S} = \mathcal{H} \times R$. Note that $\mathcal{S}$ is not user dependent; however upper semicontinuous utilities will be considered, giving an indirect way to restrict a user strategy set. For instance, if a user wants to start at a specific origin and reach a specific destination, this can be encoded in the utility function by defining it to be $-\infty$ on any route not connecting these two vertices. Moreover, $\mathcal{H}$ will be assumed to be a (bounded) interval $[h_m, h_M]$.

The travel times on an arc $a$ are modelled by an *arc travel time function* taken in $\mathcal{C}(\mathbb{R})$. If $\tau_a$ is the travel time function on arc $a$, then the quantity $\tau_a(h)$ is the time required to go through the arc when entering $a$ at $h$. Each arc is endowed with an *arc travel time model* $t_a$. A travel time model $t_a$ is a function taking as input a cumulated flow and returning an arc travel time function. Physically, an arc travel time model is simply a compact notation for traffic models. A more detailed description of travel time models is provided below.

A dynamic congestion game can be seen as a temporal extension of the congestion games introduced by Rosenthal (1973).

## Organization of the chapter

Section 1 gives the main tools and notations used. Since continuity results will be the main technical aspect of our work, we will carefully define the topologies of our different sets in this section. Section 1 also presents a theorem from Khan (Theorem 3.4), a powerful existence result on games with a continuum of users. In Section 2, we precisely describe the model we are working with. The following section – Section 4 – exposes our two main results: the consistency of our model (Proposition 3.7) and the existence theorem (Theorem 3.10). The proofs are presented in Section 5. The proof of Theorem 3.10 consists mainly of establishing that Khan's theorem can be applied.

# 1 Mathematical tools and notations

## 1.1 Sets and topologies

### 1.1.1 Measures

We use $\mathcal{M}(E)$ to denote the set of finite (Borel) measures on a metric space $E$. The measures will be denoted by capital letters, in order to be consistent with the traditional notation for dynamic user equilibrium models where *cumulated flows* are denoted by capital letters. A cumulated flow is a quantity of users on a time interval – in particular, it can be seen as a measure on time.

We will systematically use the weak convergence topology on any set of measures encountered in the chapter. A sequence of measures $M_n$ defined on a set $E$ is said to *converge weakly toward a measure $M$ on $E$* if

(i) $\limsup_{n\to+\infty} M_n(F) \leq M(F)$ for any closed subset $F$ of $E$, and

(ii) $\limsup_{n\to+\infty} M_n(E) = M(E)$.

There exists a metric $\rho$ (the Prohorov metric for instance), such that convergence for this metric is equivalent to weak convergence. It will be used in the proof of Lemma 3.14.

A set $\mathcal{M}(E)$ with the weak topology has the following property: when $E$ is a compact metric space (for instance when $E = \mathcal{H} = [h_m, h_M]$), $\mathcal{M}(E)$ is compact (Hildenbrand, 1974, page 49).

For more information about weak convergence, see (Topsoe, 1970).

### 1.1.2 Restrictions and marginals of measures

Let $M$ be a measure on a Cartesian product $A \times B$. Then $M_A$ – also called the *marginal of $M$ on $A$* – denotes the measure on $A$ such that $M_A(I) = M(I \times B)$ for each measurable subset $I \subseteq B$.

Let $M$ be a measure on $\mathbb{R}$. For any $h \in \mathbb{R}$, we denote by $M|_h$ the *restriction* of $M$ on $]-\infty, h]$, defined such that $M|_h(J) := M(J \cap ]-\infty, h])$ for all measurable subsets $J$ of $\mathbb{R}$. We extend this notation to the measure on $\mathbb{R} \times R$. If $M$ is such a measure, $M|_h(J \times R') = M((J \cap ]-\infty, h]) \times R')$ for all measurable subsets $J$ of $\mathbb{R}$ and all subsets $R'$ of $R$.

**Claim 3.1.** *If $h_2 > h_1$, then for any measure $M$, we have $M|_{h_1} = M|_{h_2}|_{h_1}$.*

The proof is straightforward.

### 1.1.3 Continuous mappings

Let $E$ and $F$ be two metric spaces, and let $\mathcal{C}(E,F)$ denote the set of all continuous maps from $E$ to $F$. When $F=\mathbb{R}$, the set $\mathcal{C}(E,F)$ is denoted, for short, by $\mathcal{C}(E)$.

When $E$ is compact, the set $\mathcal{C}(E,F)$ is endowed with the topology of uniform convergence. In particular when $E$ and $F$ are subsets of $\mathbb{R}$, it is equivalent to the topology induced by the $||\cdot||_\infty$ norm. As no topological arguments are used on sets $\mathcal{C}(E,F)$ when $E$ is not compact, such sets are not endowed with any topology.

The following lemma will be useful:

**Lemma 3.2.** *Let $I$ be a closed interval of $\mathbb{R}$ and $f:\mathcal{M}(I)\to\mathcal{C}(\mathbb{R})$ and $g:\mathcal{M}(I)\to\mathcal{C}(I)$ be two continuous functions. Then $Y\mapsto f[Y]\circ g[Y]$ is continuous.*

*Proof.* Let $\epsilon>0$ and $Y\in\mathcal{M}(I)$.

By continuity of $f$, there is an $\eta_1>0$ such that $\rho(Y,Y')\leq\eta_1$ implies $||f[Y]-f[Y']||_\infty\leq\epsilon/2$.

By uniform continuity of $f[Y]$ on the image of $g[Y]$, which is compact, there is an $\eta_2>0$ such that for all $h,h'\in\text{Im } g[Y]$, when $|h-h'|\leq\eta_2$, we have $|f[Y](h)-f[Y](h')|\leq\epsilon/2$.

By continuity of $g$, there is an $\eta_3>0$ such that $\rho(Y,Y')\leq\eta_3$ implies $||g[Y]-g[Y']||_\infty\leq\eta_2$.

Now define $\eta:=\min(\eta_1,\eta_3)$. For all $Y'\in\mathcal{M}(I)$ such that $\rho(Y,Y')\leq\eta$, we have

$$\begin{aligned}||f[Y]\circ g[Y]-f[Y']\circ g[Y']||_\infty &\leq ||f[Y]\circ g[Y]-f[Y]\circ g[Y']||_\infty \\ &\quad +||f[Y]\circ g[Y']-f[Y']\circ g[Y']||_\infty \\ &\leq \epsilon/2+\epsilon/2 \\ &\leq \epsilon.\end{aligned}$$

□

### 1.1.4 Upper semicontinuous functions

The utility of a player of dynamic congestion game given the other players strategy is modelled as a semicontinuous function. Consider a strategy set $\mathcal{S}$ and assume that $\mathcal{S}$ is a compact metric space. The function $u:E\to\bar{\mathbb{R}}=\mathbb{R}\cup\{-\infty,+\infty\}$ is said to be *upper semicontinuous* if its *hypograph* is

closed. Recall that the hypograph of a function $f : E \to \bar{\mathbb{R}}$ is given by the set $\{(x, y) \in E \times \bar{\mathbb{R}} : f(x) \geq y\}$. We denote by $\mathscr{S}_{\mathcal{S}}$ the space of upper semicontinuous functions $\mathcal{S} \to \bar{\mathbb{R}}$.

For the space $\mathscr{S}_{\mathcal{S}}$ the sup norm topology is no longer available, so we endowed it with the *hypotopology*. Hypotopology has been introduced by Dolecki, Salinetti and Roger (1983) and simply relies on the observation that that every upper semicontinuous function has a closed hypograph. Two functions are "close" if their hypographs are "close'. The chosen topology on the space of closed subsets of $\mathcal{S}$ is the closed convergence topology. It has the following valuable property: when $\mathcal{S}$ is a compact metric space, the set of all closed subsets endowed with the closed convergence topology is a compact metrizable space (see for instance Hildenbrand, 1974, page 19).

Note that $\mathcal{C}(\mathcal{S}) \subset \mathscr{S}_{\mathcal{S}}$. A natural question is then to ask if whether the hypotopology and the sup-norm topology are comparable. From Khan (1989, page 135) we have the following result: when $\mathcal{S}$ is compact, the sup-norm topology is finer than the hypotopology. That is to say that convergence in the sup-norm topology implies convergence in the hypotopology. The converse is not true.

| Notation | Definition | Topology |
|---|---|---|
| $\mathcal{M}(E)$ | The set of finite (Borel) measures on a topological space $E$ | Weak convergence topology |
| $\mathcal{C}(E,F)$ | Set of all continuous maps from $E$ to $F$ | Topology of the uniform convergence (when $E$ is compact) |
| $\mathcal{C}(\mathscr{S}_{\mathcal{S}})^a$ | Compact notation for $\mathcal{C}(\mathcal{M}(\mathcal{S}),\mathscr{S}_{\mathcal{S}})$ for a metric space E | Topology of the uniform convergence |
| $\mathscr{S}_E$ | Set of upper semicontinuous functions $E \to \bar{\mathbb{R}}$ | Hypotopology |

Table 3.1: Summary of the notations for sets ($E$ is a metric space)

[a]Will be introduced in the following section

## 1.2 Games with a continuum of users and Khan's theorem

### 1.2.1 Mas-Colell Games

One of the main approaches to games with a continuum of players was introduced by Mas-Colell (1984) as a reformulated version of Schmeidler (1973). On the basis of Hart, Hildenbrand and Kohlberg (1974), Mas-Colell represents a *game* as a probability measure $U$ on the space of utility functions $\mathcal{U}$, where $\mathcal{U}$ is defined as the space of continuous mappings from $\mathcal{S} \times \mathcal{M}(\mathcal{S})$ to $\mathbb{R}$. Given the strategy $s$ chosen by a player characterized by $u$ in $\mathcal{U}$ and the strategy distribution of all players $\boldsymbol{X} \in \mathcal{M}(\mathcal{S})$, $u(\boldsymbol{X}, s)$ is the utility enjoyed by the player[1]. A Nash equilibrium is then defined as:

**Definition 3.3.** *For a game $U$, a probability measure $D$ on $\mathcal{S} \times \mathcal{U}$ is a* Nash equilibrium *if*

1. $D_{\mathcal{U}} = U$.

[1]Note that in a Mass-Colell game players differ only by their utility functions and only the distribution of played strategies matters *i.e.* who plays what is irrelevant. Hence there are said to be *anonymous games*. This notably implies that the introduction of a space of players' names is unnecessary.

2. $D\big(\{(s,u) \in \mathcal{S} \times \mathcal{U} : u(\boldsymbol{X}, s) \geq u(\boldsymbol{X}, s') \text{ for all } s' \in \mathcal{S}\}\big) = 1$, *with* $\boldsymbol{X} := D_{\mathcal{S}}$.

Essentially, the formulation of the equilibrium states that the volume of players with a decision that is optimal relative to the overall strategy distribution is the total volume of players. Here the probability measure $D$ is to be interpreted in terms of repartition and does not imply that users act probabilistically. Definition 3.3 interprets itself easily in terms of pure strategies. Informally $D(\{s\} \times \{u\})$ simply gives the number of players characterized by $u$ playing the strategy $s$.

### 1.2.2 Khan's generalization and existence result

Mas-Colell (1984) proved the existence of an equilibrium under the assumption of $\mathcal{S}$ being a compact metric space. In an attempt to generalize this approach to upper semicontinuous utility functions, Khan proposes a slightly different model. In a Mas-Colell game, each player is characterized by a utility function $u : \mathcal{S} \times \mathcal{M}(\mathcal{S}) \to \mathbb{R}$. Khan uses an alternative view: a utility function is seen as a family of functions from $\mathcal{S}$ to $\bar{\mathbb{R}}$ parameterized by elements of $\mathcal{M}(\mathcal{S})$. In simpler terms, one can rewrite $u(\boldsymbol{X}, s) = \hat{u}[\boldsymbol{X}](s)$ in the previous definition, hence seeing a *game* as a distribution on the space of continuous mappings $\mathcal{M}(\mathcal{S}) \to \mathscr{S}_{\mathcal{S}}$. For the sake of readability, we will denote the set of such functions $\mathcal{C}(\mathscr{S}_{\mathcal{S}})$ instead of $\mathcal{C}(\mathcal{M}(\mathcal{S}), \mathscr{S}_{\mathcal{S}})$.

In Khan's extension, a Nash equilibrium is defined exactly as above, with $\mathcal{U}$ now being $\mathcal{C}(\mathscr{S}_{\mathcal{S}})$ and once we have substituted $u(\boldsymbol{X}, s)$ and $u(\boldsymbol{X}, s')$ respectively by $\hat{u}[\boldsymbol{X}](s)$ and $\hat{u}[\boldsymbol{X}](s')$. Khan showed the following theorem (Khan, 1989):

**Theorem 3.4** (Khan)**.** *Assume that the strategy set $\mathcal{S}$ is a compact metric space and let a probability measure $U \in \mathcal{M}(\mathcal{C}(\mathscr{S}_S))$ be a game. Then there exists a Nash equilibrium.*

Mas-Colell games have been applied to static user equilibrium, a successful approach which leads to important theoretical advances. They are known as congestion games in the game theory community (see for instance Milchtaich, 2005). In a congestion game players are drivers on a road network and their strategies are the possible routes on this network. The strategy distribution $\boldsymbol{X}$ hence gives the proportion of drivers choosing each route, which in transport science terminology would be the flows assigned to each route.

| Quantity | Notation | Mathematical nature |
|---|---|---|
| Set of the possible player's strategies | $\mathcal{S}$ | Non-empty compact, metric space |
| Strategy distribution of all players | $\boldsymbol{X}$ | A finite Borel measure on $\mathcal{S}$ |
| Set of strategy distributions | $\mathcal{M}(\mathcal{S})$ | Endowed with the weak convergence topology |
| Set of utility functions | $\mathcal{U}$ | $\mathcal{C}(\mathscr{S}_{\mathcal{S}})$ Endowed with the uniform convergence topology |
| Set of pay-off functions of a player given the other players' strategies | $\mathscr{S}_{\mathcal{S}}$ | The set of upper semi-continuous functions on $\mathcal{S}$ endowed with the hypotopology |
| Game | $U$ | A probability measure on $\mathcal{U}$ |

Table 3.2: Summary of the notations for Mas-Colell games (with Khan's formalism)

In the following section, the dynamic user equilibrium is formulated as a Mas-Colell game and the corresponding games are referred to as dynamic congestion games. To do so, we build a set of specific utility functions so that each of them encode the behaviour of a network user of a specific OD pair. A dynamic congestion game is defined as a measure on this latter set. It is then shown (Section 4) that these functions are continuous and consequently that dynamic congestion games are Mass-Colell games (or more precisely Khan's extensions of Mass-Colell games).

## 2 The model

In *a dynamic congestion game*, we have, on one hand, a directed graph $G = (N, A, \boldsymbol{T}_A)$ where $\boldsymbol{T}_A = (t_a)_{a\in A}$ are the travel time models associated with each arc (precisely defined below). This is the supply side. On the other

hand, we have a continuum of road users endowed with utility functions. This is the demand side.

Denoting by $R$ the set of acyclic directed paths (the set of *routes*) and by $\mathcal{H}$ a vompact time interval $\mathcal{H} \subset \mathbb{R}$, the strategy set is $\mathcal{S} := \mathcal{H} \times R$. Each user chooses his strategy from $\mathcal{S}$, that is, a departure time $h$ and a route $r = a_1, a_2, \ldots, a_n$ and goes through arcs in the order they appear in the route, entering $a_{i+1}$ as soon as he leaves arc $a_i$.

The exclusion from consideration of a sequence of arcs that does not encode a route between the origin-destination pair of a user is treated by the utility function, which will be infinitely negative for such choices.

The presentation of the model will be divided in two parts. First we introduce the network flow model: given a measure on the set of strategies representing the choices of the users, how do we compute the departure times of a user for each arc of his chosen route? Then, we define dynamic congestion games using the formalism of Definition 3.3.

## 2.1 Network flow model

Each arc $a$ is endowed with a *travel time model* $t_a : \mathcal{M}(\mathbb{R}) \to \mathcal{C}(\mathbb{R})$. If $Y$ is a measure such that for each measurable subset $J$, the quantity $Y(J)$ is the number of users having entered arc $a$ at an instant in $J$, the value $t_a[Y](h)$ is the time needed to travel along arc $a$ when travel is started at $h$. Such a measure $Y$ is called the *cumulated flow on* $a$. Consequently, the following question arises: once all users have made a choice of strategy, how do we deduce the entering instant of each user for each arc of his chosen route?

To do so, we are going to introduce the *cumulated flow function on arc* $a$, denoted $\Phi_a$, which is entirely determined by the arc travel times $t_a$. The physical meaning of $\Phi_a$ will be the following: for a distribution of user's strategy $\boldsymbol{X} \in \mathcal{M}(\mathcal{S})$, the quantity $\Phi_a[\boldsymbol{X}]$ is the cumulated flow of users on arc $a$ resulting from the propagation of the users over the network.

It remains to explicit formally each function $\Phi_a$ using the travel time models $t_a$. A useful notion is that of *arc exit time function*, $H_a : \mathcal{M}(\mathbb{R}) \to \mathcal{C}(\mathbb{R})$ which is defined by

$$H_a[Y](h) := h + t_a[Y](h) \quad \text{for } Y \in \mathcal{M}(\mathbb{R}) \text{ and } h \in \mathbb{R} \tag{3.1}$$

Given a cumulated flow $Y$ on an arc $a$, the number $H_a[Y](h)$ is the instant at which the arc $a$ is left when it has been entered at $h$. If we choose $J \subset \mathbb{R}$,

the subset $H_a[Y]^{-1}(J)$ is all the instants at which arc $a$ can be entered in order to leave it at some instant in $J$.

The arc cumulated flows $(Y_a)_{a\in A}$ can now be formally derived from a strategy distribution.

**Definition 3.5** (Dynamic network loading problem)**.** *The arc cumulated flows $(Y_a)_{a\in A}$ induced by a strategy distribution $\boldsymbol{X} \in \mathcal{M}(\mathcal{S})$ is a collection of (Borel) measures on $\mathbb{R}$ such that there exists $(Y_a^r)$ for all $r \in R$ and $a \in A$ satisfying the system:*

$$Y_a = \sum_{r\in R:\, a\in r} Y_a^r \tag{3.2}$$

*and for all $r = a_1, \ldots, a_n \in R$*

$$\begin{array}{lcll} Y_{a_1}^r & = & X^r & \text{(i)} \\ Y_{a_i}^r & = & Y_{a_{i-1}}^r \circ (H_{a_{i-1}}[Y_{a_{i-1}}])^{-1} \quad \text{for } i = 2, \ldots, n & \text{(ii)} \\ Y_a^r & = & 0 \qquad\qquad\qquad\qquad\qquad\quad \text{if } a \notin r. & \text{(iii)} \end{array} \tag{3.3}$$

*With $X^r$ the measure define by $X^r(J) := \boldsymbol{X}(J \times \{r\})$ for all measurable subsets $J$ of $\mathbb{R}$.*

*Finding the arc cumulated flows $(Y_a)_{a\in A}$ for a given strategy distribution $\boldsymbol{X} \in \mathcal{M}(\mathcal{S})$ is* ***the dynamic network loading problem****.*

We say that $Y_a^r$ is the *cumulated flow on arc $a$ with respect to route $r$* for each $J \subseteq \mathbb{R}$. Indeed the quantity $Y_a^r(J)$ is the number of users whose chosen route is $r$ and who enter arc $a$ for some instant in $J$. Similarly $X^r$ is denoted as the *cumulated flow on $r$*, which counts the number of users starting route $r$ on any interval: for each $J \subseteq \mathbb{R}$, the quantity $X^r(J)$ is the number of users who start route $r$ for some instant in $J$.

Equation (3.2) simply states that users going through arc $a$ can be decomposed by routes. Equality (i) of Equation (3.3) expresses that the number of users entering the first arc of a given route $r$ during an interval $J$ is the number of users entering the route $r$ during $J$. Equality (ii) expresses that the number of users having chosen route $r$ and entering arc $a_i$ in $J$ is equal to the number of users having chosen route $r$ and leaving arc $a_{i-1}$ in $J$ (in our model, there is no delay between the arcs). Equality (iii) expresses that if an arc $a$ does not belong to a route $r$, nobody having chosen $r$ will travel along $a$.

Until now we have no guarantee that the arc entry time and cumulated flows functions are unambiguously defined in Definition 3.5. A proposition

below (Proposition 3.7) will ensure that under five assumptions on $t_a$, existence and uniqueness of the solutions of system (3.2-3.3) are guaranteed for all $\boldsymbol{X} \in \mathcal{M}(\mathcal{S})$. In this case, the functions $\Phi_a$ can be properly defined through

$$\Phi_a[\boldsymbol{X}] := Y_a \quad \text{for all } a \in A.$$

We define another function – *the route exit time function* – that will be useful. For all $r = a_1, \ldots, a_n \in R$, let

$$H^r[\boldsymbol{X}] := H_{a_n}[Y_{a_n}] \circ H_{a_{n-1}}[Y_{a_{n-1}}] \circ \ldots \circ H_{a_1}[Y_{a_1}].$$

Note that even if the notations $H^r$ and $H_a$ are similar and their physical meanings are close, the first one depends on the whole measure on the strategies, while the second depends only on the cumulated flow on the arc.

## 2.2 The set of utility functions

Assume given a set $(t_a)_{a\in A}$ of travel time models and the corresponding arc entry and cumulated flow functions. Each user is identified by a collection of functions $u_r : \mathbb{R}^{n+1} \to \bar{\mathbb{R}}$, one for each route $r$ in $R$ (where $n$ denotes the number of arcs of route $r$). Denote by $(\mathcal{U}_r)_{r\in R}$ the set of admissible functions $u_r$. $(\mathcal{U}_r)_{r\in R}$ can be interpreted as the space of the user characteristics.

The utility function of a user characterized by $(u_r)_{r\in R}$ in the sense of Mas-Colell then comes from the following expression:

$$\hat{u}[\boldsymbol{X}](h, r = a_1, \ldots, a_n) := u_r(h_0, h_1, \ldots, h_n) \tag{3.4}$$

with $h_0 = h$ and $h_i = H^{a_1,\ldots,a_i}[\boldsymbol{X}](h)$.

$\hat{u}[\boldsymbol{X}](s)$ is the pay off of the user represented by $(u_r)_{r\in R}$ when he plays $s$ (*i.e.* when he takes the travel decision $s = (h, r)$) against the distribution $\boldsymbol{X}$ of users' strategies (*i.e.* while the decisions of the other users are summarized by $\boldsymbol{X}$). $h_0, h_1, \ldots, h_{n-1}$ are the instants at which the arcs $a_1, a_2, \ldots, a_n$ are entered by the user, and $h_n$ is the instant at which he leaves the last arc. Equation (3.4) expresses that the utility depends not only on the time to complete the whole routes, but also on the instants at which the arcs have been entered. Such a feature enables to represent a large number of interesting situations. For instance there is an increasing interest for time-varying tolling policies and DUE models are typically used to assess such schemes (see Aguiléra and Wagner, 2009, or Chapter 10). This could also accounts

for short intermediate stops that are only possible within a restricted time window *e.g.* picking up dry cleaning before the outlet closes.

Note that at this point, we have no idea on the mathematical properties of $\hat{u}[\boldsymbol{X}]$ except it is a map $\mathcal{S} \mapsto \bar{\mathbb{R}}$. In the following, we will introduce assumptions on the travel time models $t_a$ and on the route utility functions $u_r$ and prove they imply $\hat{u}[\boldsymbol{X}] \in \mathscr{S}_{\mathcal{S}}$. It will also be shown that $\hat{u} \in \mathcal{C}(\mathcal{S})$.

The set of utility functions $\hat{u}$ is denoted $\mathcal{U}\left[(\mathcal{U}_r)_{r\in R}, (t_a)_{a\in A}\right]$. We are now ready to set the dynamic congestion game definition:

**Definition 3.6.** *A dynamic congestion game is a probability measure $U$ over $\mathcal{U}\left[(\mathcal{U}_r)_{r\in R}, (t_a)_{a\in A}\right]$.*

From definition 3.6, it is possible to define the Nash equilibrium of a dynamic congestion games in a similar fashion as for Mas-Colell's games (see definition 3.3).

# 3 A simple example of dynamic congestion game

Up to now the concept of dynamic congestion game is rather abstract. We detour briefly to present a simple illustration of the basic idea: the two-routes problem with heterogeneous users w.r.t. their value of time. This case study is classical in the transport economics literature (*e.g.* Verhoef and Small, 1999) and has been studied extensively for static congestion. In a few words, users are allowed to choose between two routes: one is slow but has a low toll while the other is faster but more expansive. As users value time savings differently, what is the resulting equilibrium?

## 3.1 Presentation

The considered network is shown in Figure 3.1. It is composed of two arcs, $a_1$ and $a_2$, and has just one origin destination pair, o-d, connected by two routes, $r_1 = a_1$ and $r_2 = a_2$. Both routes are priced with a flat toll, denoted respectively $p_1$ and $p_2$. Arc $a_1$ and $a_2$ have an exit capacity of $k$ and a free flow travel time of $t_0$ that we will set to 0 for the sake of simplicity. Each arc is endowed with the corresponding bottleneck travel time model (see Chapter 1). The presentation of how we formally define the bottleneck

model on the space of cumulated flows $\mathcal{M}(\mathbb{R})$ is delayed until Chapter 4. For now, we just consider cumulated flows $Y$ that admit a density $y$ and define the bottleneck model by the standard relation:

$$\dot{t}_{a_i}[Y](h) = \begin{cases} \dfrac{y(h) - k_i}{k_i} & \text{if } t_{a_i}[Y](h) > t_{0,a_i} \text{ or } y(h) - k_i > 0 \\ 0 & \text{otherwise} \end{cases}, \tag{3.5}$$

where $\dot{()}$ is the differentiation w.r.t. to $h$. Note that, in this case, the dynamic network loading problem presented in (3.3) is straightforward and that we have $X^{r_i} = Y_i$ for $i \in \{1, 2\}$.

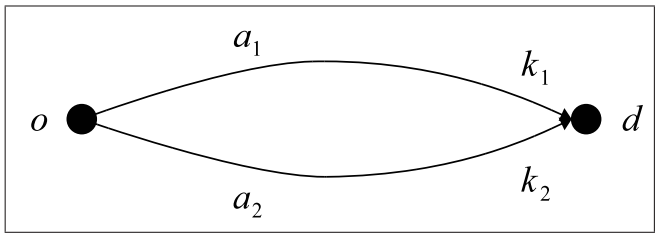



Figure 3.1: A simple network

Now assume a set of users wish to go from $o$ and $d$. Each user is characterized by two parameters $h \in \mathcal{H}$, his departure time, and $\nu \in [\nu_m, \nu_M]$, his value of time. A user chooses the route $r_i$ that minimizes his generalized cost $\nu t[Y_i](h) + p_i$. Let us introduce the spaces $\mathcal{U}_{r_1}$ and $\mathcal{U}_{r_2}$ of functions $u_r^{(h,\nu)}$ of the type:

$$u_{r_i}^{(h,\nu)}(h_0, h_1) = \begin{cases} -\big(\nu.(h_1 - h_0) + p_i\big) & \text{if } h_0 = h \\ -\infty & \text{otherwise} \end{cases} \tag{3.6}$$

One might wonder if these sets of utility functions allow us to encode the assumed users' behaviour on our small network. As utility is reduced to the general cost of transport and that the departure times other than $h$ are forbid to the user by the second line of (3.6), the route utility functions $u_{r1}^{(h,\nu)}$,$u_{r2}^{(h,\nu)}$ correctly represents the behaviour of a user characterized by $(h, \nu)$.

Having defined $\mathcal{U}_{r_1}$ and $\mathcal{U}_{r_2}$ as well as $t_{a_1} = t_{a_2}$, the global setting of the dynamic congestion game is in place and a game is simply represented by a distribution[2] on $\mathcal{U}[(\mathcal{U}_r)_{r\in\{r_1,r_2\}}, (t_a)_{a\in\{a_1,a_2\}}]$. This distribution can be identified to a distribution on the space of user characteristics $\mathcal{C} = \mathcal{H} \times [\nu_m, \nu_M]$.

[2]Naturally, not all the possible distributions on $\mathcal{U}\left[(\mathcal{U}_r)_{r\in\{r_1,r_2\}}, (t_a)_{a\in\{a_1,a_2\}}\right]$ corre-

## 3.2 Analytical resolution for a uniform distribution

Consider a given game denoted $U$ and understood as a distribution over $\mathcal{C} = \mathcal{H} \times [\nu_m, \nu_M]$. Assume it is a uniform distribution of density $\mu$. This subsection exposes how to compute an equilibrium of $U$. Our approach is the following. We assume there exists an equilibrium $D$, a measure on $\mathcal{S} \times \mathcal{C}$ and derive the necessary conditions it must satisfy.

Denote $Y_1(J) := D_{\mathcal{S}}(J \times \{r_1\})$ and $Y_2 := D_{\mathcal{S}}(J \times \{r_2\})$ the corresponding cumulated flows on arc $a_1$ and $a_2$, respectively, and $\Delta t[Y_1, Y_2](h) := t[Y_1](h) - t[Y_2](h)$. As $D$ is an equilibrium, a user $(h, \nu)$ must be assigned to his optimal route $r$ given by the following rule:

$$r = 1 \quad \text{if } \nu > \frac{p_2 - p_1}{\Delta t[Y_1, Y_2](h)}$$

$$r = 2 \quad \text{if } \nu < \frac{p_2 - p_1}{\Delta t[Y_1, Y_2](h)}$$

This leads us to introduce the quantity $\nu^\star(h) := \dfrac{p_2 - p_1}{\Delta t[Y_1, Y_2](h)}$ for any $h \in \mathcal{H}$. The quantity $\nu^\star(h)$ is the critical value of time dividing $[\nu_m, \nu_M]$ into two sets of value of times, $[\nu_m, \nu^\star(h)]$ and $[\nu^\star(h), \nu_M]$ representing respectively the users patronizing route 1 and 2. When $\nu^\star(h) \in [\nu_m, \nu_M]$, this yields the following relationships on $Y_1$ and $Y_2$:

$$Y_1([0,h]) = \int_0^h \int_{\nu_m}^{\nu^\star(h)} \mu \; d\nu dh = \mu \int_0^h (\nu^\star(h) - \nu_m) dh$$

$$Y_2([0,h]) = \int_0^h \int_{\nu^\star(h)}^{\nu_M} \mu \; d\nu dh = \mu \int_0^h (\nu_M - \nu^\star(h)) dh$$

By differentiating the previous equations, the densities of $Y_1$ and $Y_2$, denoted respectively $y_1$ and $y_2$ can be expressed relatively to $\nu^\star(h)$:

$$y_1(h) = \mu.(\nu^\star(h) - \nu_m) \quad \text{and} \quad y_2(h) = \mu.(\nu_M - \nu^\star(h)) \tag{3.7}$$

Now assume that both $t_{a_1}[Y_1]$ and $t_{a_2}[Y_2]$ are congested travel times *i.e.* that they satisfy the first equation of (3.5) for almost every $h \in \mathcal{H}$. Combining (3.7) and (3.5) and replacing $\nu^\star(h)$ by its expression w.r.t. $\Delta t[Y_1, y_2](h)$,

spond to the situation presented above. Indeed we wish that a user has the same value of time and departure time on both routes. Thus the considered distribution $U$ should be such that the measure of the set of utility functions build from pairs of route utility functions with different values of times and/or departure times is zero.

yields:

$$\begin{aligned}&\frac{k_1 k_2}{\mu}\dot{\Delta}t[Y_1, Y_2](h)\\ &\quad = (k_1+k_2)\mu\frac{(p_1-p_2)}{\Delta t[Y_1, Y_2](h)} + k_2\nu_M - k_1\nu_m\end{aligned} \tag{3.8}$$

Equation (3.8) is an explicit differential equation in $\Delta t[Y_1, Y_2]$ and it thus admits a unique solution satisfying the initial condition $\Delta t[Y_1, Y_2](0) = t_{0,a_1} - t_{0,a_2}$. Once (3.8) is solved the cumulated flows $Y_1$ and $Y_2$ can be immediately derived from the expressions a few lines above. The equilibrium distribution $D$ can also be established as follows:

For any $E \subset \mathcal{C}$ and $J \subset \mathcal{H}$,
$D(J \times r_1 \times E) := U(E \cap \{(h,\nu) : h \leq v^\star(h) \text{ and } h \in J\}$
For any $E \subset \mathcal{C}$ and $J \subset \mathcal{H}$,
$D(J \times r_2 \times E) := U(E \cap \{(h,\nu) : h \geq v^\star(h) \text{ and } h \in J\}$

Recall that this reasoning is only valid if the resulting arc travel time functions $t_{a_1}[Y_1]$ and $t_{a_2}[Y_2]$ are congested on $\mathcal{H}$, *i.e.* if $t_{a_i}[Y_i] > t_{0,a_i}$ almost everywhere on $\mathcal{H}$. Treating the general case would require to consider a set of differential equations, one for each congested and uncongested periods and to iteratively resolve them. This case is not treated here but a very similar situation can be found in Chapter 5.

The analytical resolution of (3.8) is tedious and requires the use of non elementary functions, namely product log functions. However it can be solved numerically. The results for the parameters of Table 3.3 are shown in Figure 3.2. The interpretation is very simple: at first the value of $\nu^\star(0)$ is exactly 8 €/h so the users are evenly dispatched among the two routes. As the capacity of arc $a_1$ is higher than the one of arc $a_2$, at first the difference of travel times is decreasing. As time goes, route $r_1$ becomes less attractive for the users with high values of time and the flow on route $r_1$ decreases while the one on route $r_2$ increases. This results in a decrease in $\dot{\Delta}t[Y_1, Y_2]$ and after 0.2 hour the system reaches a steady state where the flows on each route as well as the difference of travel time between each route is constant.

| $k_1$ (pcu/h) | $k_2$(pcu/h) | $p_2 - p_1$ (€) | |
|---|---|---|---|
| 2000 | 1000 | 4 | |
| $t_{0,a_1} - t_{0,a_2}$ (h) | $\mathcal{H}$ | $[\nu_m, \nu_M]$(€/h) | $\mu$ |
| 0.5 | $[6, 10]$ | 800 | 0.5 |

Table 3.3: Numerical parameters for the illustration

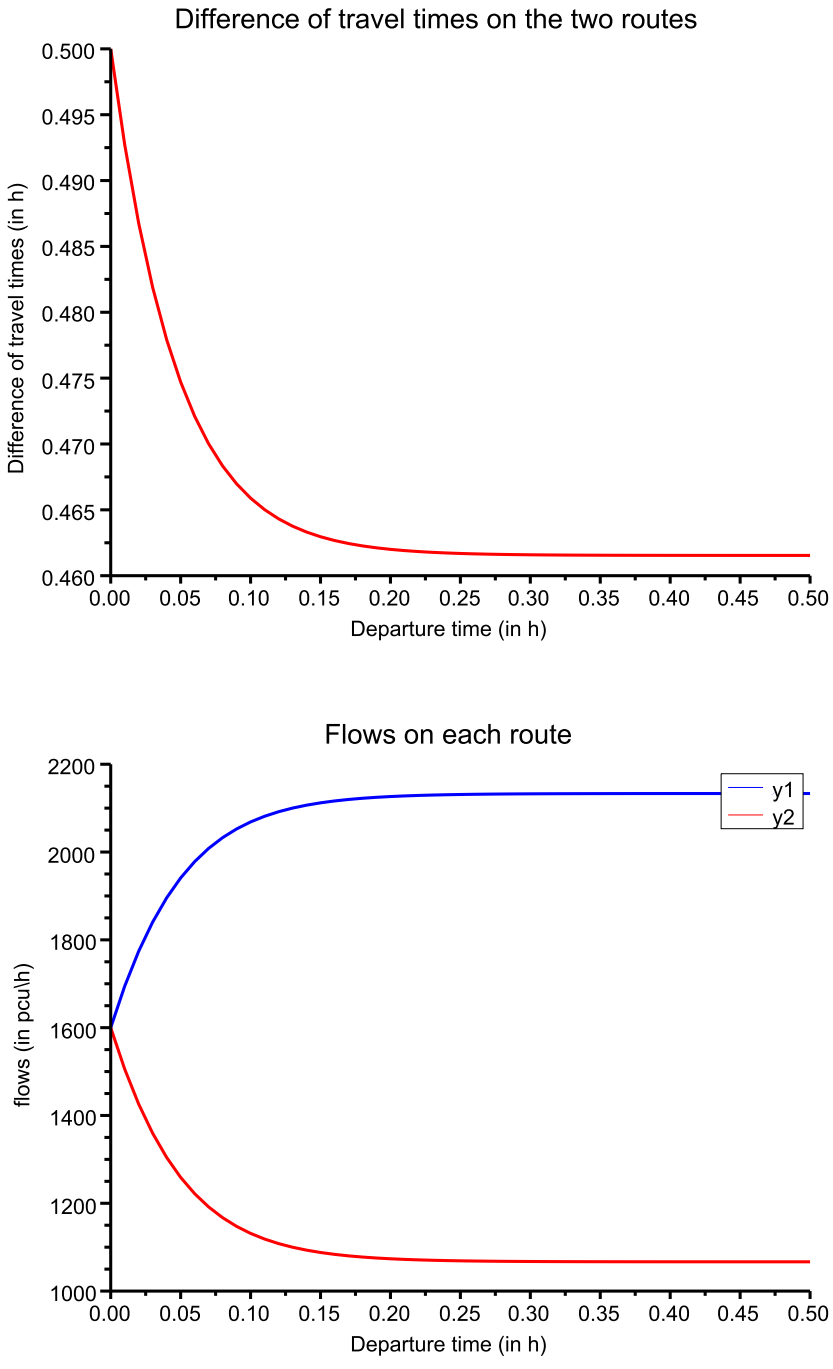



Figure 3.2: Equilibrium of a simple dynamic congestion game

# 4 Main results

## 4.1 Existence and uniqueness of the solution of the dynamic network loading problem

To establish the existence and uniqueness of the solutions to the dynamic network loading problem as defined in Subsection 2.1 and hence show that the functions $\Phi_a$ are well-defined, we need to introduce five (natural) assumptions on the nature of the travel time models $t_a$.

**Assumption I.** [Continuity] $t_a : \mathcal{M}(\mathbb{R}) \to \mathcal{C}(\mathbb{R})$ *is continuous.*

**Assumption II.** [No infinite speed] *There exists* $t_{\min} > 0$ *such that for all* $Y \in \mathcal{M}(\mathbb{R})$ *and all* $h \in \mathbb{R}$*, we have* $t_a[Y](h) > t_{\min}$.

**Assumption III.** [Finiteness] *There exists a continuous map* $t_{\max} : \mathbb{R}_+ \to \mathbb{R}_+$ *such that* $t_a[Y](h) \leq t_{\max}(Y(\mathbb{R}))$ *for all* $h \in \mathbb{R}$.

**Assumption IV.** [Strict FIFO] *Let* $Y \in \mathcal{M}(\mathbb{R})$*. The map* $h \mapsto h + t_a[Y](h)$ *is non decreasing. Moreover, for* $h_1 < h_2$ *in* $\mathbb{R}$ *such that* $Y[h_1, h_2] \neq 0$*, we have* $h_1 + t_a[Y](h_1) < h_2 + t_a[Y](h_2)$.

**Assumption V.** [Causality] *For all* $h \in \mathbb{R}$ *and* $Y \in \mathcal{M}(\mathbb{R})$*, we have* $t_a[Y|_h](h) = t_a[Y](h)$.

Assumption I states that a small variation on the cumulated flow on an arc leads to a small variation of the arc travel time function. Assumption II amounts to say that the time needed to travel along an arc is bounded from below. The finiteness condition (Assumption III) assumes that if we wait for a sufficiently long time, there will be no user left on any arc. The FIFO condition (Assumption IV) states that if two users enter an arc in a given order, they leave it in the same order. Finally, Assumption V implies that the arc travel time depends on the users that have already entered this arc, but not on the ones that will.

We then have:

**Proposition 3.7.** *Given a strategy distribution* $\boldsymbol{X} \in \mathcal{M}(\mathcal{S})$*, system (3.3) has a unique solution* $(Y_a)_{a \in A}$*. Moreover* $\boldsymbol{X} \mapsto Y_a$ *is continuous for each* $a \in A$*. Hence* $\Phi_a : \boldsymbol{X} \in \mathcal{M}(\mathcal{S}) \mapsto Y_a \in \mathcal{M}(\mathbb{R})$ *is well-defined and continuous for all* $a \in A$.

**Corollary 3.8.** $H^r$ *is well-defined and continuous for all* $r \in R$.

## 4.2 Existence result

In order to establish the existence result we will focus on a specific category of utility functions. A general existence result on dynamic congestion games remains an open problem.

**Definition 3.9** (Route utility with departure time penalty)**.** *A route utility function $u_r$ incorporates a departure time penalty if there exists an upper semicontinuous function $p_r \in \mathscr{S}(\mathcal{H})$ and a function $\tau_r \in \mathcal{C}(\mathbb{R}^{n+1})$ such that:*

$$u_r(h_0, \dots, h_n) = p_r(h_0) + \tau_r(h_0, h_1, \dots, h_n)$$

$p_r$ can be interpreted as a departure time penalty. In the context of transport science, the departure time penalty is a standard modeling feature (see Heydecker and Addison, 2005). For instance during evening commutes, travelers might be unable to leave before the end of the work period. Assuming $p_r$ upper semicontinuity is particularly desirable as it allows to forbid certain departure times for specific users by arbitrary setting $p_r$ to $-\infty$ on open subsets of $\mathcal{H}$. The main reason of Definition 3.9 is technical as keeping the general definition of utility functions introduced in Subsection 2.2 makes it difficult to compose them with route exit time functions while keeping well behaved functions. Indeed whereas we have Lemma 3.2 for continuous functions, there is no equivalent for upper semicontinuous ones.

Note that it is still possible to impose forms of penalty at arrival, by encompassing it in $\tau_r$. However, those penalties will continuously vary with the arrival time. Although it would be of clear interest to relax this assumption, it is important to remind that most of the transport models in fact assume continuous penalty at arrival.

**Theorem 3.10.** *Given a set of arc travel time functions $(t_a)_{a\in A}$ satisfying Assumptions I-V and compact sets $(\mathcal{U}_r)_{r\in R}$ of route utility functions with departure time penalty, every measure $U$ on $\mathcal{U}\left[(\mathcal{U}_r)_{r\in R}, (t_a)_{a\in A}\right]$) admits a Nash equilibrium distribution.*

# 5 Proofs

## 5.1 Proof of Proposition 3.7

We first introduce:

1. for an arc $a$, $\boldsymbol{Y}_a$ is the collection $(Y_a^r)_{r\in R}$. It can alternatively be seen as an element of $\mathcal{M}(\mathbb{R}\times R)$;

2. for an arc $a$, the function $\psi_a$ from $\mathcal{M}(\mathbb{R}\times R)$ to itself, defined as follows for each measurable subset $J$ of $\mathbb{R}$. For any $r$ let:

$$\psi_a^r(\boldsymbol{Y}_a)(J) := \begin{cases} Y_a^r(H_a(Y_a)^{-1}(J)) & \text{if } a\in r \\ 0 & \text{if not,} \end{cases}$$

with $Y_a = \sum_{r\in R:\, a\in r} Y_a^r$ (Equation (3.2)). Then $\psi_a(\boldsymbol{Y}_a) := (\psi_a^r(\boldsymbol{Y}_a))_{r\in R}$, which we also see as an element of $\mathcal{M}(\mathbb{R}\times R)$.

Note that the fact that $\psi_a^r(\boldsymbol{Y}_a)$ is a measure is a consequence of the continuity of $H_a[Y_a]$ (Assumption I).

$\psi_a$ can be interpreted as a kind of "transfer function", which, given a cumulated flow on arc $a$ – that is a distribution of users entering arc $a$ – returns a distribution of users leaving arc $a$, and this, decomposed for each route $r$ containing arc $a$.

Let us first state two lemmas regarding $\psi_a$ properties, used in the proof of Lemma 3.13, which in turn is used in the proof of Proposition 3.7.

**Lemma 3.11.** *Consider a bounded interval $I'\subseteq\mathbb{R}$ and an arc $a$. Then $\psi_a$ is continuous on the set of measures having their support in $I'$.*

*Proof.* Take a converging sequence $\boldsymbol{Y}_n \to \boldsymbol{Y}$ of measures on $\mathbb{R}\times R$ having their support in $I'$ and define as usual $Y_n := \sum_{r\in R:\, a\in r} Y_n^r$ and $Y := \sum_{r\in R:\, a\in r} Y^r$.

Consider $f_n := H_a[Y_n]$ and $f := H_a[Y]$. Note that the sequence $||f_n - f||_\infty$ converges to 0 by continuity of $H_a$.

Choose $r\in R$. We want to prove that $\limsup_n Y_n^r(f_n^{-1}(J)) \le Y^r(f^{-1}(J))$ for any interval $J\subseteq\mathbb{R}$ with equality when $J=\mathbb{R}$. This latter case (that is when $J=\mathbb{R}$) is straightforward since $Y_n^r(f_n^{-1}(\mathbb{R})) = Y_n^r(\mathbb{R}) \to Y^r(\mathbb{R})$ when $n$ goes to infinity.

Take an interval $J=[h_1,h_2]$ in $\mathbb{R}$. The interval $J$ can be assumed to be bounded since all measures are assumed to have support in $I'$ and since the Assumption III is satisfied.

We can assume that $Y_n^r(f_n^{-1}(J)) \neq 0$ for an infinite sequence of $n$, otherwise there is nothing to prove.

Define now

$$\begin{array}{llll} h_{1,n} := \inf f_n^{-1}(J), & & h_{2,n} := \sup f_n^{-1}(J) \\ \tilde{h_1} := \inf f^{-1}(J), & & \tilde{h_2} := \sup f^{-1}(J) \\ h_1^* := \liminf_n h_{1,n}, & & h_2^* := \limsup_n h_{2,n} \end{array}$$

We have

$$h_1 \leq f_n(h_{1,n}) - f(h_{1,n}) + f(h_{1,n}).$$

Hence, using the fact that $f$ is an increasing function we get

$$f(h_1^*) \geq h_1.$$

Similarly, we get

$$f(h_2^*) \leq h_2,$$

and thus

$$h_2 \geq f(h_2^*) \geq f(h_1^*) \geq h_1.$$

Hence we have:

$$\tilde{h_1} \leq h_1^* \text{ and } \tilde{h_2} \geq h_2^* \tag{3.9}$$

Let us now prove the following result: *Let $\epsilon > 0$. There exists $h_1' < h_1^*$ and $h_2' > h_2^*$ such that $Y^r([h_1', h_2']) \leq Y^r([h_1^*, h_2^*]) + \epsilon$.*

Take a sequence of closed intervals $(I_n)$ converging to $[h_1^*, h_2^*]$ such that $[h_1^*, h_2^*]$ is strictly included in $I_n$ for any $n$. According to the sequential continuity of measures (Hildenbrand, 1974, page 43) $\lim_n Y^r(I_n) = Y^r(\lim_n I_n) = Y^r([h_1^*, h_2^*])$, so there exists $n'$ such that $Y^r(I_{n'}) \leq Y^r([h_1^*, h_2^*]) + \epsilon$. Take $[h_1', h_2'] = I_{n'}$.

Now, for $n$ big enough, we have $f_n^{-1}(J) \subseteq [h_1', h_2']$. Hence, for $n$ big enough

$$\begin{array}{lll} Y_n^r(f_n^{-1}(J)) \leq Y_n^r([h_1', h_2']) & \text{(by monotonicity of a measure)} \\ \leq Y^r([h_1', h_2']) + \epsilon & (Y_n^r \text{ converges to } Y^r) \\ \leq Y^r([h_1^*, h_2^*]) + 2\epsilon & \\ \leq Y^r([\tilde{h_1}, \tilde{h_2}]) + 2\epsilon & \text{(according to (3.9)).} \end{array}$$

□

**Lemma 3.12.** *For all $h \in \mathbb{R}$ and $\boldsymbol{Y}_a \in \mathcal{M}(\mathbb{R})$, we have*

$$\psi_a(\boldsymbol{Y}_a)|_{H_a[Y_a](h)} = \psi_a(\boldsymbol{Y}_a|_h)|_{H_a[Y_a|h](h)} \tag{3.10}$$

*and*

$$\psi_a(\boldsymbol{Y}_a)|_{h+t_{\min}} = \psi_a(\boldsymbol{Y}_a|_h)|_{h+t_{\min}}. \tag{3.11}$$

Let us interpret Equation (3.11) – Equation (3.10) is only a step in the proof of Equation (3.11). It means that the distribution of users leaving arc $a$ before $h+t_{\min}$ depends only on the distribution of users entering arc $a$ before $h$. Recall that $t_{\min}$ is a lower bound on the time needed to travel along an arc $a$.

*Proof of Lemma 3.12.* As soon as the first equality is true, the second one is also true, as a consequence of Claim 3.1 and of Assumption II.

Let us prove the first equality. Fix $h \in \mathbb{R}$, $Y, Y' \in \mathcal{M}(\mathbb{R})$ and $E$ a measurable subset of $\mathbb{R}$. We first prove two properties.

*Property 1*: $H_a[Y|_h]^{-1}(E) \cap ]-\infty, h] = H_a[Y]^{-1}(E) \cap ]-\infty, h]$.

Indeed, for $h' \leq h$, we have $H_a[Y|_h](h') = H_a[Y|_h|_{h'}](h') = H_a[Y|_{h'}](h') = H_a[Y](h')$ with the help of Claim 3.1 for the second equality and of Assumption V for the first and third equalities.

*Property 2*: If $E \subseteq ]-\infty, H_a[Y](h)]$, and if $Y' \leq Y$, then $Y'\,(H_a[Y]^{-1}(E) \cap ]h, +\infty[) = 0$.

Indeed, let $h' \in H_a[Y]^{-1}(E) \cap ]h, +\infty[$. We have $h' > h$ and $H_a(Y)(h') \leq H_a[Y](h)$. According to Assumption IV, we have then $Y[h, h'] = 0$, and hence $Y'[h, h'] = 0$.

Take now $h \in \mathbb{R}$, $\boldsymbol{Y}_a \in \mathcal{M}(R \times \mathbb{R})$, $r \in R$ and $J$ a measurable subset of $\mathbb{R}$. Define $E := J \cap ]-\infty, H_a[Y_a](h)]$. The set $E$ is a measurable subset of $\mathbb{R}$ and it is such that $E \subseteq ]-\infty, H_a[Y_a](h)]$. Note that $Y_a^r \leq Y_a$ when $a \in r$.

$$
\begin{array}{rcll}
\psi_a^r(\boldsymbol{Y}_a)|_{H_a[Y_a](h)}(J) & = & Y_a^r(H_a[Y_a]^{-1}(E)) & \text{(by definition)} \\
& = & Y_a^r(H_a[Y_a]^{-1}(E) \cap ]-\infty, h]) & \\
& & \quad + Y_a^r(H_a[Y_a]^{-1}(E) \cap ]h, +\infty[) & \text{(since } Y_a^r \text{ is a measure)} \\
& = & Y_a^r(H_a[Y_a]^{-1}(E) \cap ]-\infty, h]) & \text{(according to Property 2)} \\
& = & Y_a^r(H_a[Y_a|_h]^{-1}(E) \cap ]-\infty, h]) & \text{(according to Property 1)} \\
& = & Y_a^r|_h(H_a[Y_a|_h]^{-1}(E)) & \text{(by definition of } |_h) \\
& = & Y_a^r|_h(H_a[Y_a|_h]^{-1}(E)) & \text{(since } Y_a^r|_h \text{ is a measure)} \\
& = & \psi_a^r(\boldsymbol{Y}_a|_h)|_{H_a[Y_a|_h](h)}(J) & \text{(by definition).}
\end{array}
$$

□

The following lemma states how the user strategies $\boldsymbol{X}$ induce the cumulated flows $(Y_a^r)_{a \in A, r \in R}$ on each arc $a$ with respect to each route $r$.

**Lemma 3.13.** *Fix $k \in \mathbb{N}$. Given a measure $\boldsymbol{X} \in \mathcal{M}(\mathcal{S})$ (a strategy distribution), there exists a unique collection $(Y_a^r)_{a\in A, r\in R}$ of elements of $\mathcal{M}(\mathbb{R})$ such that for all routes $r = a_1, a_2, \ldots, a_n$*

$$(E_k) \quad \begin{cases} Y_{a_1}^r &= X^r|_{kt_{\min}} \\ Y_{a_i}^r &= \psi_{a_{i-1}}^r(\boldsymbol{Y}_{a_{i-1}})\Big|_{kt_{\min}} & \text{for } i = 2, \ldots, n \\ Y_a^r &= 0 & \text{if } a \notin r \end{cases}$$

*Moreover, for any $a$, the map $\Phi_a^k : \boldsymbol{X} \mapsto Y_a := \sum_{r:\, a\in r} Y_a^r$, where $(Y_a^r)_{a\in A, r\in R}$ is the solution of $(E_k)$, is continuous.*

Informally, Lemma 3.13 says that it is possible to construct a sequence of measures on $\mathcal{S}$, with each of its elements representing the users' progress over their routes, with a time step of $t_{\min}$. The proof of the lemma relies on Assumption II (an arc can not be traversed at an infinite speed), which highlights the crucial importance of this assumption in our approach.

*Proof of Lemma 3.13.* The proof works by induction on $k$. For $k = 0$, define $Y_a^r := 0$ for all $r$ and $a$. $(Y_a^r)_{a\in A, r\in R}$ is a solution of $E_k$, which gives the existence part of Lemma 3.13 for $k = 0$. Uniqueness is straightforward.

Suppose now that $k \geq 0$ and that we have proved the lemma till $k$.

*Existence and continuity*: Let $(\boldsymbol{Y}_a)_{a\in A} = (Y_a'^r)_{a\in A, r\in R}$ be the solution of $(E_k)$. We want to prove that $(E_{k+1})$ has a solution. Define $(Y_a^r)_{a\in A, r\in R}$ for all routes $r = a_1, a_2, \ldots, a_n$ by

$$\begin{array}{rcll} Y_{a_1}^r &:=& X^r|_{(k+1)t_{\min}} & \\ Y_{a_i}^r &:=& \psi_{a_{i-1}}^r(\boldsymbol{Y}'_{a_{i-1}})\Big|_{(k+1)t_{\min}} & \text{for } i = 2, \ldots, n \\ Y_a^r &:=& 0 & \text{if } a \notin r \end{array}$$

According to this definition and Lemma 3.11, $\boldsymbol{Y}_a$ depends continuously on $\boldsymbol{X}$.

Note that, according to Claim 3.1, we have then for all $a \in A$, $r \in R$

$$Y_a'^r = Y_a^r|_{kt_{\min}} \tag{3.12}$$

We check that the collection $(Y_a^r)$ is solution of $(E_{k+1})$. The first and the last equalities of $(E_{k+1})$ are straightforward. Let us check the second one. Let $r = a_1, a_2, \ldots, a_n$ be a route in $R$.

$$\begin{aligned} Y^r_{a_i} &= \psi^r_{a_{i-1}}\left(\boldsymbol{Y}'_{a_{i-1}}\right)\Big|_{(k+1)t_{\min}} && \text{(by definition of } \boldsymbol{Y}_{a_i}) \\ &= \psi^r_{a_{i-1}}\left(\boldsymbol{Y}_{a_{i-1}}\big|_{kt_{\min}}\right)\Big|_{(k+1)t_{\min}} && \text{(according to Equation (3.12))} \\ &= \psi^r_{a'}\left(\boldsymbol{Y}_{a'}\right)\big|_{(k+1)t_{\min}} && \text{(according to Equation (3.11) of Lemma 3.12)} \end{aligned}$$

*Uniqueness*: Assume that we have two collections $(\boldsymbol{Y}_a)_{a\in A}$ and $(\boldsymbol{Z}_a)_{a\in A}$ solutions of $(E_{k+1})$. Yet, $(\boldsymbol{Y}_a|_{kt_{\min}})_{a\in A}$ and $(\boldsymbol{Z}_a|_{kt_{\min}})_{a\in A}$ are solutions of $(E_k)$. Hence, by induction,

$$(\boldsymbol{Y}_a|_{kt_{\min}})_{a\in A} = \left(\boldsymbol{Z}_a|_{kt_{\min}}\right)_{a\in A} \tag{3.13}$$

We can write the chain of equalities for any $a \in A$

$$\begin{aligned} Y^r_a &= \psi^r_{a'}\left(\boldsymbol{Y}_{a'}\right)\big|_{(k+1)t_{\min}} && \text{(since } \boldsymbol{Y}_a \text{ is solution of } (E_{k+1})) \\ &= \psi^r_{a'}\left(\boldsymbol{Y}_{a'}\big|_{kt_{\min}}\right)\Big|_{(k+1)t_{\min}} && \text{(according to Equation (3.11) of Lemma 3.12)} \\ &= \psi^r_{a'}\left(\boldsymbol{Z}_{a'}\big|_{kt_{\min}}\right)\Big|_{(k+1)t_{\min}} && \text{(according to Equation (3.13))} \\ &= Z^r_a && \text{(since } \boldsymbol{Z}_a \text{ is solution of } (E_{k+1})). \end{aligned}$$

□

We are now in position to prove Proposition 3.7.

*Proof of Proposition 3.7.* Recall that $\boldsymbol{X}(\mathcal{H}\times R) = \sum_{r\in R} X^r(\mathcal{H}) = 1$. Let $\tau := \max_{x\in[0,1]} t_{\max}(x)$. According to Assumption III, for any route $r = a_1, a_2, \ldots, a_n$, a direct induction on $i$ leads to $\boldsymbol{Y}_{a_i} = \boldsymbol{Y}_{a_i}|_{i\tau}$ (no one leaves arc $a_i$ after $i\tau$). Hence, any cumulated flows $(\boldsymbol{Y}_a)_{a\in A}$ solution of (3.3) is solution of Equation $(E_k)$ for a large enough $k$. It means that for a $k$ large enough, we have $\Phi^k_a = \Phi_a$. Existence, continuity, and uniqueness are consequences of Lemma 3.13. □

It remains to prove Corollary 3.8.

*Proof of Corollary 3.8.* Since we have by definition

$$H^{a_1,\ldots,a_n}[\boldsymbol{X}] = H_{a_n}[\Phi_{a_n}[\boldsymbol{X}]] \circ \ldots \circ H_{a_1}(\Phi_{a_1}[\boldsymbol{X}]),$$

the corollary is a straightforward consequence of Proposition 3.7 and Lemma 3.2. □

## 5.2 Proof of the main theorem

*Proof of Theorem 3.10.* Theorem 3.10 is a direct consequence of the following lemma (Lemma 3.14) and of Theorem 3.4. □

**Lemma 3.14.** *The set of utility functions* $\mathcal{U}\left[(\mathcal{U}_r)_{r\in R}, (t_a)_{a\in A}\right]$ *considered in a dynamic congestion game is a measurable subset of* $\mathcal{C}(\mathscr{S}_{\mathcal{S}})$.

*Proof.* For any $r = a_1, \ldots, a_n$ denote $\phi_r[\boldsymbol{X}](h) := (h, H^{a_1}[\boldsymbol{X}](h), H^{a_1,a_2}[\boldsymbol{X}](h), \ldots, H^{a_1,\ldots,a_n}[\boldsymbol{X}](h))$.

To prove that $\boldsymbol{X} \mapsto \hat{u}[\boldsymbol{X}]$ is continuous, it is enough to prove that $\boldsymbol{X} \mapsto u_r \circ \phi_r[\boldsymbol{X}]$ is continuous for any $r$ (see Equation (3.4)). Yet, according to Corollary 3.8, $\phi_r$ is continuous. Lemma 3.2 applied to $\tau_r \circ \phi_r(\boldsymbol{X})$ and the fact that the sup-norm topology is finer than the hypotopology implies that $\boldsymbol{X} \mapsto u_r \circ \phi_r[\boldsymbol{X}]$ is continuous.

Finally, the measurability comes from the fact that for each $r$, the map

$$\mathcal{J} : (p_r, \tau_r) \mapsto (\boldsymbol{X} \mapsto p_r + \tau_r \circ \phi_r[\boldsymbol{X}])$$

is continuous (since $\phi_r[X]$ is continuous) and $\mathcal{U}_r$ is compact. Indeed, $\mathcal{U}\left[(\mathcal{U}_r)_{r\in R}, (t_a)_{a\in A}\right]$ is then the image of a compact set by the continuous map $\mathcal{J}$. □

# Conclusion

This chapter introduces dynamic congestion games as a general framework for the dynamic user equilibrium problem, bringing new results into the transport field from the field of mathematical economics. It is shown that the existence of a Nash equilibrium in dynamic congestion games is guaranteed under five natural assumptions on the arc travel time models. This was achieved by studying the property of the dynamic network loading problem and showing it is well posed *i.e.* that it admits a unique solution and that the resulting map between the route cumulated flows and the arc cumulated flows is continuous. As this latter proof is constructive, a numerical algorithm can naturally be derived for the dynamic loading problem. This is the object of the Appendix D.

It is important to note the wide range of modelling possibilities that dynamic congestion games offer. Correctly defining the set of route utility functions allows an incredibly large set of variations. For instance, utilities that varies non linearly with travel time might be considered. Road pricing

strategies can be embedded in the utility functions by adding maluses on specific routes, possibly only for specific types of users. As the route utility is expressed as a function of the time of entrance on every arc of the route, those pricing schemes might be time-varying. Finally the possibility of intermediate stops from which the user might derive some utility, typically short shopping stops, might be represented. As far as congestion modelling is concerned, the assumptions we considered seems *a priori* weak and it is reasonable to think that they include a wide range of specific travel time models.

Although dynamic congestion games allow to represent most of the classic physical features of dynamic transport models, the question of the formal equivalence between this formulation of the user equilibrium and standard ones has yet to be examined. This is the topic of the next chapter.

Chapter 4

# Application to the dynamic traffic assignment problem on a network of bottlenecks

The existence result of Chapter 3 is fairly general and notably apply to most of the problems of dynamic equilibrium assignment. Those models, although commonly used in practice for transport planning, lack theoretical foundations and results of existence have been established only in very restrictive cases.

In Chapter 2, the most common dynamic assignment problem, the so-called dynamic Wardrop assignment problem is presented. We begin this chapter by recalling the standard formulation of this problem, with a particular emphasis on the difference with the formalism of dynamic congestion games (Section 1). In Section 2, it is shown the dynamic Wardrop assignment problem can be written as a dynamic congestion game and thus has a solution under the general arc travel time assumptions we stated in the previous Chapter. In the last section (Section 3), a common travel time model of the transport literature, the bottleneck model, is reviewed and it is shown that it is a well-behaved travel time model in the sense stated in the previous chapter.

## 1 Dynamic Wardrop assignment

The simplest assignment model can be formulated as follows. Consider a travel demand, described by flows of users between each origin-destination pair, and assume each user is allowed to choose his travel route, but not his departure time. We study the possible assignments of traffic flows to routes

connecting each origin-destination pair. The question is the following: is there an assignment such that no route is assigned with a non-zero flow of vehicles at a time $h$ if there are routes with smaller travel times? Such an assignment is said to satisfy the *dynamic Wardrop principle*. Note that the terminology in the transport literature varies from one author to another, and that what we call dynamic Wardrop assignment is also termed user equilibrium assignment.

A formal statement of the dynamic assignment problem is presented below. Before doing so, let us raise a few comments on the mathematical nature of traffic flows in transport models other than ours. Existing models represent vehicle flows by integrable functions, whereas our formulation is based on measures on a bounded interval $\mathcal{H}$ (route flows) or on a larger interval $\bar{\mathcal{H}}$ (arc flows), so it is useful to associate each element of $L_1(\mathcal{H}, \mathbb{R}_+)$, the sets of positive integrable functions on $\mathcal{H}$ with an element of $\mathcal{M}(\mathcal{H})$. Thus to a flow $x \in L_1(\mathcal{H}, \mathbb{R}_+)$ we associate the cumulated flow $X$ defined by $X(]-\infty, h]) = \int_{-\infty}^{h} x(h')dh'$.

Given a map $x$ in $L^1(\mathbb{R}, \mathbb{R}_+)$, we can associate with $x$ a measure $X$ on $\mathbb{R}$ defined by $X(]-\infty, h]) = \int_{-\infty}^{h} m(h')dh'$. If $X$ can be written in this form, it is said to *have a density*. Physically if $M$ is a cumulated flow, $m$ is the corresponding instantaneous flow.

A measure $X$ is said to be *absolutely continuous* with respect to $X'$ if $X(A) = 0$ for every set $A$ for which $X'(A) = 0$. In finite-dimensional spaces, the absolutely continuous measures with respect to the Lebesgue measure are exactly the ones that have a density.

Recall that in this case $X$ is an absolutely continuous measure (see Chapter 3, Section 1). To guarantee a unique mapping between an absolutely continuous measure and a measurable function, the functions in $L_1(\mathcal{H}, \mathbb{R}_+)$ equal almost everywhere are quotiented out[1]. A similar operation is performed on $L_1(\bar{\mathcal{H}}, \mathbb{R}_+)$

We can now formulate the dynamic Wardrop assignment problem. Consider a dynamic transport network $G = (N, A, \boldsymbol{T}_A)$, with arc travel time models $\boldsymbol{T}_A = (t_a)_{a \in A}$ and an *origin-destination matrix* $(x_{od})_{o \in N, d \in N}$, each element of the matrix being a function in $L_1(\mathcal{H}, \mathbb{R}_+)$. The arc travel time models are defined from $L_1(\bar{\mathcal{H}}, \mathbb{R}_+)$ to $\mathcal{C}(\mathbb{R}_+)$, using the identification exposed in the paragraph above, from the set of absolutely continuous measures on $\bar{\mathcal{H}}$

[1] In other terms we consider the set obtained by identifying the elements $f$ and $g$ such that $f$ equals to $g$ almost everywhere.

to $\mathcal{C}(\bar{\mathcal{H}}, \mathbb{R}_+)$. These arc travel time models are thus restrictions on $L_1(\bar{\mathcal{H}}, \mathbb{R}_+)$ of the arc travel time models as they are defined in the Section 2 of Chapter 3. Hence we call them *restricted arc travel time models*. We will see below (Subsection 2.1) how to extend them to the whole set $\mathcal{M}(\bar{\mathcal{H}})$.

An *assignment of the traffic* is an element $\boldsymbol{x} = (x^r)_{r\in R}$ of $L_1(\mathcal{H}, \mathbb{R}_+)^R$ such that $\sum_{r\in R_{od}} x^r(h) = x_{od}(h)$ for all $(o,d) \in N \times N$ and $h \in \mathcal{H}$, with $R_{od}$ denoting the set of routes connecting $o$ to $d$. For each route $r = a_1, \ldots, a_n$, let us define a route travel time model $t_r$ by

$$t_r[\boldsymbol{X}](h) := H^r[\boldsymbol{X}](h) - h, \tag{4.1}$$

where $\boldsymbol{X}$ is the measure whose density is $\boldsymbol{x}$. The quantity $t_r[\boldsymbol{X}](h)$ is then the time needed to traverse route $r$ when leaving at instant $h$ for an assignment $\boldsymbol{x}$. Note that $t_r$ is well defined as long as $\Phi_{a_i}[\boldsymbol{X}]$ returns an absolutely continuous measure when $\boldsymbol{X}$ is one. Indeed, $H^r[\boldsymbol{X}]$ is defined by the expression $H_{a_n}[\Phi_{a_n}[\boldsymbol{X}]] \circ \ldots \circ H_{a_1}[\Phi_{a_1}[\boldsymbol{X}]]$ with $r = a_1, \ldots, a_n$. Under a sixth assumption on the arc travel time functions $t_a$ (Assumption VI), we will see that this condition can be satisfied.

We have intentionally used the same notations for the users' strategy distributions and the traffic assignments, as it is natural to interpret $X^r$ as a cumulated flow of vehicles ; $X^r(]-\infty; h])$ counts the number of vehicles that have already entered route $r$.

**Definition 4.1** (Dynamic Wardrop Assignment Problem)**.** *Find an assignment* $\boldsymbol{x} \in L_1(\mathcal{H}, \mathbb{R}_+)^R$ *such that whenever* $r, r' \in R_{od}$

$$x^r(h) > 0 \Rightarrow t_r[\boldsymbol{X}](h) \leq t_{r'}[\boldsymbol{X}](h), \text{ for almost every } h \in \mathcal{H}$$

Note that at each instant the flow of vehicles leaving an origin is fixed, *i.e.* vehicles can not adjust their departure time. Without loss of generality, it is assumed below that $\sum_{od\in N\times N} \int_{\mathcal{H}} x_{od}(h) = 1$.

# 2 An existence result for the dynamic Wardrop assignment problem

Assumptions I-V have been stated for standard travel time models not for *restricted* ones. Now they can straightforwardly be adapted to restricted

travel time models. We claim that if the restricted travel time models satisfy Assumptions I-V, as well as an additional one, the so-called bounded variations assumption, there exists an equilibrium assignment.

**Assumption VI.** [Bounded variations] *There is a real number $K$ such that for any absolutely continuous measure $Y_a$ with derivative $y_a$, the map $h \mapsto t_a[Y_a](h)$ is differentiable almost everywhere (a.e.) on $\mathbb{R}$ and $h \mapsto \frac{1}{y_a(h)}.\frac{dt_a[Y_a]}{dh}(h)$ is smaller than or equal to $K$ a.e. .*

Assumption VI is slightly less intuitive than the ones in Section 4 of Chapter 3, but for traffic propagation it makes physical sense. Intuitively, if a flow of vehicles $x$ enters an arc $a$ then the outgoing flow would be something like $x/\left(1+\frac{dt_a(x)}{dh}\right)$. Consequently, Assumption VI implies that if the inflow on an arc is bounded by a constant $K$, then the outflow is bounded by a constant $K'$ that depends only on $K$ and $t_a$. In other words, Assumption VI ensures that for a given assignment problem, there is a bound on the flows on each arc of the network (*i.e.* on the density of $\Phi_a[\boldsymbol{X}]$) so that for any traffic assignment unreasonably high flows of traffic will not be observed.

Let us construct a game from the assignment problem. We have already started with our choice of notations, but a few issues still need to be addressed. First, we have to extend the definition of the restricted arc travel time functions, as they are still only defined on $L_1(\bar{\mathcal{H}}, \mathbb{R}_+)$ – this is the purpose of Subsection 2.1. Then, using an adequate set of utility functions (Subsection 2.2), Theorem 3.10 of Chapter 3 will tell us that there is an equilibrium, but this equilibrium is a measure $D$ leading to an assignment $X^r := D_S(\mathcal{H} \times \{r\})$ that might not have a derivative in $L_1(\mathcal{H}, \mathbb{R}_+)$. Our equilibrium might not be an equilibrium is the sense above. This last issue is the object of Lemma 4.3 in Subsection 2.3.

## 2.1 Extension of the travel time models

Consider $K$ a constant and denote $\mathcal{M}^{\leq K}(\mathbb{R})$ the set of positive measurable functions essentially bounded by $K$, *i.e.* measures $M$ such that for any interval $J$, one has $M(J) \leq K\mu(J)$ (where $\mu$ is the usual Lebesgue measure on $\mathbb{R}$). It is easy to see that these measures are absolutely continuous and that their set is closed. As $t_a$ is continuous on $\mathcal{M}^{\leq K}(\mathbb{R})$, there exists a continuous extension $t'_a$ of $t_a$ on $\mathcal{M}(\mathbb{R})$ by the Tietze-Dugundji extension theorem (Dugundji, 1951). Since we can require the extension to remain

within the convex hull of the arc travel times functions (defined for absolutely continuous measures) whose derivative are bounded a.e. by $K$, we will have continuous extensions that will satisfy Assumptions I-III and the first sentence of Assumption IV. To enforce the satisfaction of the last two assumptions, we make the following redefinition:

$$t_a[Y](h) := t'_a[Y|_h](h) + \int_0^h \rho\left(Y|_{h'}, \mathcal{M}^{\leq K}(\mathbb{R})\right) dh',$$

where $\rho$ is the Prohorov metric.

Hence, for any constant $K$, we can construct well defined arc travel time models (for a given assignment problem) that extend the restricted arc travel time models. By abuse of notation, they will also be denoted $t_a$. Note that although the extension depends on $K$, we have omitted any explicit reference to it.

Finally, it remains to choose a constant $K$. What would be an appropriate value for $K$? It should be high enough such that no equilibrium assignment induces flows $Y_a = \Phi_a[\boldsymbol{X}]$ such that $y_a \geq K$ on a non null measurable set of $\mathbb{R}$. Assumption VI guarantees the existence of such a constant.

## 2.2 Utility functions

We can now build the game associated with the dynamic Wardrop assignment problem. Consider a distribution of users $U$ on the set $\mathcal{RC}$ (which stands for "route choice") of continuous utility functions of the following type

$$\hat{u}_{h^*,od}[\boldsymbol{X}](h,r) = \begin{cases} -t_r[\boldsymbol{X}](h) & \text{if } h = h^* \text{ and } r \in R_{od}, \\ -\infty & \text{otherwise.} \end{cases} \tag{4.2}$$

Here, $t_r$ is defined on any measure $\boldsymbol{X} \in \mathcal{M}(\mathcal{S})$. It is the extension of the $t_r$ defined by Equation (4.1) when we use the extension of the arc travel time $t_a$ in Subsection 2.1 (and hence the extension of the $H_a$ and $H^r$).

The interpretation is straightforward: each user is characterized by a departure time $h^*$ he will always prefer, and an origin-destination pair $od$ on which he will always travel. The utility of a travel decision is limited to the travel time on the route.

Denote $\tilde{t}_r$ the map defined by

$$\tilde{t}_{r,h^*}(h_0,\ldots,h_n) = \begin{cases} h_n - h_0 & \text{if } h_0 = h^* \text{ and } r \in R_{od}, \\ +\infty & \text{otherwise.} \end{cases} \tag{4.3}$$

for any $r = a_1, \ldots, a_n \in R$ and any $h^* \in \mathcal{H}$. Then taking $\mathcal{U}_r := \{-\tilde{t}_{r,h^*} : h^* \in \mathcal{H}\} \cup \{+\infty\}$, $\mathcal{RC}$ is obviously a measurable subset of $\mathcal{U}\left[(\mathcal{U}_r)_{r\in R}, (t_a)_{a\in A}\right]$. Hence we are in the framework defined in the previous chapter (Subsection 2.2 of Chapter 3).

## 2.3 Properties of the equilibrium distribution.

The set $\mathcal{R}C$ can be identified with $N \times N \times \mathcal{H}$ and according to the context a measure on $\mathcal{R}C$ is seen either as a measure on $\mathcal{C}(\mathscr{S}_\mathcal{S})$, or as a collection of measures $(U_{od})_{o\in N, d\in N}$ on $\mathcal{H}$. The latter point of view is of particular interest because of the following proposition:

**Proposition 4.2.** *If $U$ is a measure on $\mathcal{RC}$ seen as a collection $(U_{od})_{o\in N,d\in N}$, the equilibrium assignment $\boldsymbol{X}$ satisfies:*

$$U_{od} = \sum_{r\in R_{od}} X^r \tag{4.4}$$

*Proof.* Consider a measure $U$ on $\mathcal{RC}$ such that $U(\mathcal{C}(\mathscr{S}_\mathcal{S})) = U\{\hat{u}_{h,od} : h \in \mathcal{H}, od \in N \times N\}$. Let $D$ be an associated Nash equilibrium. Recall that $\boldsymbol{X} := D_\mathcal{S}$ and $U = D_{\mathcal{C}(\mathscr{S}_\mathcal{S})}$. Then for all measurable subsets $E$ of $\mathcal{H}$:

$\boldsymbol{X}(R_{od} \times E) =$

| | | |
|---|---|---|
| $=$ | $D_\mathcal{S}(R_{od} \times E)$ | (by definition of $\boldsymbol{X}$ ) |
| $=$ | $D(R_{od} \times E \times \mathcal{C}(\mathscr{S}_\mathcal{S}))$ | (by definition of a margin) |
| $=$ | $D(R_{od} \times E \times \mathcal{RC})$ | ($U$ is a measure on $\mathcal{RC}$) |
| $=$ | $D(R_{od} \times E \times \{\hat{u}_{h^*,od}$ such that $h^* \in E\})$ | (D is an equilibrium measure) |
| $=$ | $D(\mathcal{S} \times \{\hat{u}_{h^*,od}$ such that $h^* \in E\})$ | (idem) |
| $=$ | $U(\{\hat{u}_{h^*,od}$ such that $h^* \in E\})$ | (idem) |
| $=$ | $U_{od}(E)$ | (identifying $\mathcal{RC}$ with $N \times N \times \mathcal{H}$) |

□

Proposition 4.2 simply restates in measure terms that an assignment is a decomposition of these flows over the set of the routes. The following lemma is an important consequence.

**Lemma 4.3.** *Let $U$ be a measure on $\mathcal{RC}$, seen as measure on $\mathcal{C}(\mathscr{S}_\mathcal{S})$. If $U$ is absolutely continuous, every equilibrium assignment $\boldsymbol{X}$ is also absolutely continuous.*

*Proof.* Consider an absolutely continuous measure $U$ on $\mathcal{C}(\mathscr{S}_{\mathcal{S}})$ such that $U(\mathcal{C}(\mathscr{S}_{\mathcal{S}})) =$

$U(\{\hat{u}_{h,od} : h \in \mathcal{H}, od \in N \times N\})$, let $D$ be an associated Nash equilibrium and let $\boldsymbol{X} := D_{\mathcal{S}}$. According to Proposition 4.2:

$$U_{od} = \sum_{r \in R_{od}} X^r$$

Then if we have $E$ a measurable subset of $\mathcal{H}$ such that $U_{od}(E) = 0$, for all $r \in R_{od}$ we have $X^r(E) = 0$. Thus, absolute continuity of $U$ implies absolute continuity of $\boldsymbol{X}$. □

## 2.4 Theorem

We can now state the main result of the section:

**Theorem 4.4.** *Given a dynamic transport network $G = (N, A, \boldsymbol{T}_A)$ whose arc travel time function satisfy Assumptions I-VI and given an origin-destination matrix, there is a Wardrop assignment.*

*Proof.* Assume we are given an origin-destination matrix $(x_{od})$. Define $U$ a measure on $\mathcal{C}\big(\mathscr{S}_{\mathcal{S}}\big)$ such that

- $U(\mathcal{C}(\mathscr{S}_{\mathcal{S}})) = U(\mathcal{RC}) = 1$ and
- for a given pair $od \in N \times N$ and any measurable subset $J \subseteq \mathcal{H}$, we have $U(\{\hat{u}_{h,od} : h \in J\}) = \int_{h \in J} x_{od}(h)dh$.

We have just encoded our origin-destination matrix as a measure on the set of users. Note that $U$ is absolutely continuous.

According to Theorem 3.10 of Chapter 3, there exists a Nash equilibrium $D$, and according to Lemma 4.3 the equilibrium assignment $\boldsymbol{X} := D_{\mathcal{S}}$ is absolutely continuous with respect to the Lebesgue measure. Hence $\boldsymbol{X}$ admits a density, which we will denote $\boldsymbol{x}$. Recall that the cumulated flows induced by $\boldsymbol{x}$ (*i.e.* the $\Phi_a[\boldsymbol{X}]$) are essentially bounded by the constant $K$ set at the end of Subsection 2.1. Consequently we are in the part of $\mathcal{M}(\mathbb{R})$ on which the restricted arc travel time functions were originally defined.

Let $h \in \mathcal{H}$ and take any route $r$ such that $x^r(h) > 0$. Let $od$ be the origin-destination pair connected by $r$. The proof proceeds in two steps. First, we show that whenever $x^r$ is continuous in $h$, $x^r(h) > 0 \Rightarrow t_r[X](h) \leq t_{r'}[\boldsymbol{X}](h)$

for all $r' \in R_{od}$. Then, we show that this inequality holds almost everywhere.

**First step.** Let $h \in \mathcal{H}$ be such that $s$ is continuous in $h$. Now, take any route $r$ such that $x^r(h) > 0$. Let $od$ be the origin-destination pair connected by $r$. For all $\epsilon > 0$, we have $X^r([h-\epsilon; h+\epsilon]) > 0$, which can be rewritten $D(\{r\} \times \{h'\} \times \hat{u}_{h',od} : h' \in [h-\epsilon, h+\epsilon]) > 0$. Therefore, we know that for all $\epsilon > 0$, there is $h' \in [h-\epsilon, h+\epsilon]$ such that $\hat{u}_{h',od}[\boldsymbol{X}](r', h'') \leq \hat{u}_{h',od}[\boldsymbol{X}](h', r)$ for all $h'' \in \mathcal{H}$ and $r' \in R$, or, directly in terms of route travel times:

$$\begin{aligned}&\text{for all } \epsilon > 0, \text{ there is } h' \in [h-\epsilon, h+\epsilon] \text{ such that}\\ &\qquad t_{r'}[\boldsymbol{X}](h') \geq t_r[\boldsymbol{X}](h') \text{ for all } r' \in R_{od}.\end{aligned}$$

By continuity of $h \mapsto t_r[\boldsymbol{X}](h)$, we get the required inequality.

**Second step.** For a given $r$ consider $E$ the set of points such that $x^r(h) > 0$ and $t_r[\boldsymbol{X}](h) > t_{r'}[\boldsymbol{X}](h)$ for a $r'$ on the same origin-destination pair as $r$. From the previous paragraph $x^r$ is discontinuous at every $h \in E$. $E$ is measurable since $t_r$, $t_{r'}$ and $x_r$ also are. Now assume $\mu(E) = \epsilon \neq 0$, denoting $\mu$ the Lebesgue measure on $\mathbb{R}$. Then, $x^r$ being measurable, there exists a set $F$ such that the measure of its complementary $\mu(F^c) < \epsilon/2$ and $x^r$ is continuous in every $h \in K$ (Lusin Theorem (Lusin, 1912)). So $E \subseteq F^c$, a contradiction. Hence $\mu(E) = 0$.

Thus, the required inequality is valid almost everywhere.

□

# 3 Formal properties of the punctual bottleneck model with time-varying capacity

The bottleneck model has already been encountered in Chapter 1. The objectives of this section are twofold. First, the bottleneck model is generalized to the case where the exit capacity is time-varying. Second its formal properties are studied; more specifically its continuity and the satisfaction of Assumptions (I-VI) are examined.

The formalization of the bottleneck model retained in this section is a restricted travel time model in the sense exposed in Section 1. It takes as input $Y$, an absolutely continuous measure on $\bar{\mathcal{H}}$ (or equivalently a function $y \in L_1(\bar{\mathcal{H}}, \mathbb{R})$) and returns a continuous travel time function $t[Y] : \bar{\mathcal{H}} \to \mathbb{R}_+$.

## 3.1 Statement of the bottleneck model with non-time-varying capacity

From the seminal work of Vickrey there has been various statements of the bottleneck model. We present a simple intuitive one below, widely used by the transport science community (see for instance Arnott et al., 1998).

Denote $k$ the capacity of the bottleneck and consider $Y$ a cumulated volume on $\mathcal{H}$ with density $y$ and $t_0$ its free flow travel time.

Let us first define $Q[Y]$, the stock of users waiting in the bottleneck by the following equation:

$$\dot{Q}[Y](h) = \begin{cases} y(h) - k & \text{if } Q(h) > 0 \text{ or } y > k \\ 0 & \text{otherwise} \end{cases}, \tag{4.5}$$

where $\dot{()}$ is the differentiation w.r.t. to $h$. So that Equation (4.5) defines $Q[Y]$ on $\mathbb{R}$ rather than $\mathcal{H}$, extend $y$ on $\mathbb{R}$ by letting $y(h) = 0$ for $h \notin \mathcal{H}$.

The interpretation of (4.5) is straightforward: when the entrance flows exceed capacity users began to accumulate in a punctual queue until a sufficient drop in demand allows to clear all the stock of traffic. Then the travel $t[Y]$ is simply given by:

$$t[Y](h) = t_0 + \frac{Q[Y \circ H_0]}{k} \tag{4.6}$$

where $H_0 := \mathrm{id}_{\bar{\mathcal{H}}} + t_0$. The relation between $t[Y]$ and $Q$ expresses that a user arriving at $h$ has to wait for all the users already in the queue when he arrived have left before going through the bottleneck. It reflects a first in first out discipline. The use of the quantity $Y \circ H_0$ rather than $Y$ simply express that the bottleneck is located at the arc's exit.

## 3.2 Statement of bottleneck model with time-varying capacity and free flow travel time

We introduce a variant to the former model, that represents a more general case. In this formulation both capacity and free flow travel time are functions of the time. In term of assumptions and formulation, this is no new approach. A similar model can be found in (Smith and Wisten, 1995) for instance. Two equivalent formulations are proposed.

**Integral form.** Let $Y$ be a cumulated flow of users, $y$ its density, $h \mapsto k(h)$ a function of the time representing the capacity of the bottleneck at each instant, and $h \mapsto t_0(h)$ the time-varying free flow travel time. Denote $K = \int_0^h k(h)dh$. It is assumed that $k > 0$ on $\mathbb{R}$ and thus $K$ is strictly increasing. Moreover $t_0$ is assumed to be continuous, differentiable almost everywhere and such that $\dot{t}_0 > -1$ and $t_{0,\min} < t_0(h) < t_{0,\max}$. In this case, the stock evolution equation, Equation (4.5), becomes:

$$\dot{Q}[Y](h) = \begin{cases} y(h) - k(h) & \text{if } Q[Y](h) > 0 \text{ or } y(h) > k \\ 0 & \text{otherwise} \end{cases} \tag{4.7}$$

Then $t[Y](h)$ is the solution of the following equation:

$$K\left(h + t[Y](h)\right) - K \circ H_0(h) = Q[Y \circ H_0^{-1}] \circ H_0(h) \tag{4.8}$$

where $H_0 := \mathrm{id}_{\bar{\mathcal{H}}} + t_0$. Equation (4.8) is simply a reformulation of Equation (4.6) for a time-varying capacity. It states that the travel time $t[Y](h)$ can be found by integrating $k$ from $h$ until the value of the integral is equal to the stock *i.e.* until all the users waiting at $h$ have been served. Note that Equations (4.8) and (4.7) define a system in $t[Y](h)$, that can be solved chronologically for all $h \in \mathcal{H}$.

**Differential form.** For a given entrance flow $Y$, the travel time function $t[Y]$ is a continuous functions of $h$. Thus sets $\{h : t(h) = 0\}$ [resp. $\{h : t(h) = 0\}$] are countable unions of closed [resp. open] intervals. We refer to those intervals as *unqueued* [resp. *queued*] periods. We denote $Q_1 = ]q_0, q_1[, Q_2 = [q_1, q_2], \ldots, Q_{2n+1}$ the sequence of unqueued and queued periods, $q_{2k}$ and $q_{2k+1}$ being transition instants from an unqueued period to the next queued period, and from queued to unqueued, respectively. The travel time function $t[Y]$ satisfies the following equations on queued and unqueued periods.

On any queued interval the derivative of $Q[Y \circ H_0^{-1}] \circ H_0$ w.r.t. time is $\dot{H}_0(h).(y(h) - k \circ H_0(h))$, so differentiating Equation (4.8) yields:

$$y(h).\dot{H}_0(h) = k\big(h + t[Y](h)\big).\big(1 + \dot{t}[Y](h)\big) \tag{4.9}$$

On any unqueued interval, by definition:

$$y(h) \leq k(h) \text{ and } t[Y](h) = t_0(h) \tag{4.10}$$

We claim that Equations (4.9) and (4.10) are sufficient to uniquely define $t[Y]$.

**Proposition 4.5.** *Consider a continuous travel time function* $t[Y] : \bar{\mathcal{H}} \mapsto \mathbb{R}_+$ *such that* $t[Y](h) \geq t_0(h)$ *for all* $h$ *and* $t[Y](h_m) = t_0(h_m)$*. The function* $t[Y]$ *is a solution to Equation* (4.6) *if and only if there exists a sequence of instants* $(q_i)_{i=0..2n+1}$*, such as* $t[Y]$ *is a solution to Equation* (4.9) *on any* $Q_{2i+1} = [q_{2i}, q_{2i+1}]$ *and to Equation* (4.10) *on any* $Q_{2i} = [q_{2i-1}, q_{2i}]$*.*

*Proof of Proposition 4.5.* We already demonstrated the "only if" part. For the "only if" it is sufficient to consider a function $t[Y]$ as described in Proposition 4.5 and integrate it iteratively on the intervals $Q_{2i}$ and $Q_{2i+1}$ to show that $t[Y]$ is a solution to Equation (4.6). □

## 3.3 Continuity

**Topologies.** Before treating of continuity, recall the topologies endowed with the space of flows (*i.e.* the set of absolutely continuous measures on $\mathcal{H}$) and the set of travel time functions (*i.e.* $\mathcal{C}(\bar{\mathcal{H}}, \mathbb{R}_+)$). The set of travel times is endowed with topology of the uniform convergence. The topology on the set of flows is defined with respect to the cumulated flows rather than the instantaneous flows. Formally it is the weak convergence topology. Yet the weak convergence of $Y_n$ toward $Y$ is equivalent to the pointwise convergence of the cumulative distribution function of $Y_n$ (*i.e.* $h \mapsto Y_n]-\infty, h]$) toward the cumulative distribution function of $Y$ (*i.e.* $h \mapsto Y]-\infty, h]$). This result is known as the Portmanteau theorem on the convergence of measures (see Billingsley, 1995, pp 327). Now pointwise convergence of a sequence of increasing continuous functions toward a continuous function implies uniform convergence. The topology on the set of cumulated flows is thus the one induced by the following norm: $||Y||_\infty$ is the uniform norm of the cumulative distribution function of $Y$.

**A continuity statement.** We then have the following proposition:

**Proposition 4.6.** *For any capacity* $k : \mathbb{R} \to [k_{\min}, +\infty[$ *and continuous free flow travel time function* $t_0$*, the bottleneck travel time model is continuous.*

*Proof of Proposition 4.6.* Let us consider the case where $t_0 = 0$. The result can straightforwardly be extended to the case where $t_0 \neq 0$.

Consider $\eta > 0$ and $Y_1$ and $Y_2$ such that $||Y_1 - Y_2||_\infty < \eta$. By abuse of notation, we write $Y_i(h)$ for $Y_i(]\infty, h])$. We are going to show that for every $\epsilon > 0$ we can choose $\eta$ such that $||t[Y_1] - t[Y_2]||_\infty < \epsilon$.

For every $h \in \mathbb{R}$, define $q^1(h)$ and $q^2(h)$ as $q^i(h) := \max\{h' : h' \leq h \text{ and } t[Y_i](h') = 0\}$. For a given $h$ assume w.l.o.g. that $q^2(h) > q^1(h)$. Let us first remark that:

$$\begin{aligned}|K\big(q^2(h)\big) - K(q^1(h)\big) - Y_2\big(q^2(h)\big) + Y_2\big(q^1(h)\big)| \\ < Y_1\big(q^2(h)\big) - Y_1\big(q^1(h)\big) - Y_2\big(q^2(h)\big) - Y_2\big(q^1(h)\big) \\ < 2.\eta\end{aligned}$$

Then, using Equation (4.8):

$$K\big(h+t[Y_1](h)\big) - K\big(h+t[Y_2](h)\big) = \big(Y_1(h) - K(q^1(h))\big) - \big(Y_2(h) - K(q^2(h))\big)$$

So:

$$\begin{aligned}|K\big(h + t[Y_1](h)\big) - K\big(h + t[Y_2](h)\big)| < \big|Y_1(h) - Y_1(q^1(h)) - Y_2(h) + Y_2(q^1(h))\big| \\ + |K(q^2(h)) - K(q^1(h)) - Y_2(q^2(h)) + Y_2(q^1(h))|\end{aligned}$$

$$\begin{aligned}\Rightarrow |h + t[Y_1](h) - h - t[Y_2](h)|.k_{min} < |(Y_1(h) - Y_2(h)) - (Y_1(q^1(h) - Y_2(q^1(h))| \\ +2.\eta\end{aligned}$$

$$\Rightarrow |t[Y_1](h) - t[Y_2](h)| < \frac{4\eta}{k_{\min}}$$

Taking $\eta = \dfrac{k_{min}\epsilon}{4}$ leads to the conclusion.

□

## 3.4 Satisfaction of the Assumptions

The previous subsection showed that the bottleneck travel time model satisfied Assumption I, continuity. Assumptions II to VI still need to be examined. Consider a bottleneck model travel time with the assumptions of Proposition 4.6. Then:

- Assumption II, No infinite speed. As for any $Y$, $t[Y] \geq t_0 \geq t_{0,\min}$, it is straightforward.
- Assumption III, Finiteness. Taking $t_{\max}(Y(\mathbb{R})) = t_{0,\max} + \frac{Y(\mathbb{R})}{k_{\min}}$ is enough.

- Assumption IV Strict fifoness. Consider $h_1 < h_2$ in $\mathbb{R}$ such that $Y[h_1, h_2] \geq 0$. Then there exists a non null subset $I \subseteq [h_1, h_2]$ where $y > 0$. From Equation (4.9), $\dot{t}[Y](h) > -1$ during a queued period, whereas $\dot{t}[Y] = 0$ during an unqueued period. The the result follows directly.

- Assumption V, Causality. This is straightforward from the specification of the arc travel time model.

- Assumption VI, Bounded variation. The assumption is expressed by Equation (4.9).

A consequence is that there always exists a dynamic Wardrop assignment on a network of bottlenecks. This result is already known and is due to Mounce (2006).

# 4 Conclusion

This previous chapter introduces dynamic congestion games as a general framework for the user equilibrium problem. It is shown that the problem is well-posed in the sense that the existence of a Nash equilibrium is guaranteed. In this chapter, as an illustration, we exposed how to exploit this result to prove the existence of a dynamic Wardrop assignment. This result could be extended without much efforts to incorporate tolls or utilities that varies non linearly with travel time.

The application of Theorem 3.10 of Chapter 3 to prove Theorem 4.4 is quite instructive. The main difficulty lies in adapting Assumption I to commonly used travel time models. Indeed in bottleneck models, as well as in delay-volume models, when a sequence of arc incoming flows converges toward a Dirac function, the resulting travel times converge toward a discontinuous function. Hence if we try to extend straightforwardly classical travel times on the whole set of measures, this extension won't satisfy Assumption I. Thus a less rough extension was introduced so that Theorem 3.10 of Chapter 3 can be applied. Under the assumptions of Theorem 4.4, it was then verified that the equilibrium obtained by Theorem 3.10 is also a Wardrop assignment in the sense of Theorem 4.4. However, it seems unlikely that Assumption I could be alleviated as travel times need to be continuous to correctly formulate the game. For further existence results based on

Theorem 3.10, the strategy we adopted in the last section of the chapter will certainly be useful.

A possible future work could be to formulate the user equilibrium model proposed in (Lindsey, 2004) as a dynamic congestion game in an attempt to generalize Lindsey's result to a network.

Finally it remains to address the problem of the uniqueness and the stability of the equilibrium. Although the game theory standard toolbox has been helpful in this matter for the static case (see for instance (Milchtaich, 2005)), to the author's opinion the framework proposed here is too general to deal with those two issues.

# Part III

# Analytical resolutions of simple games

# Part introduction

## Objectives and structure

The previous part presented a general framework to model Dynamic User Equilibrium, the so-called dynamic congestion games. This part deals with two instances of DUE models that can be seen as simple dynamic congestion games. The focus here is on the modelling of departure time choice and user heterogeneity. In both games, the network is either a one-arc network (Chapter 5) or a two-arc network (Chapter 6) with a bottleneck travel time model. Each game is focused on a specific user characteristic that is continuously distributed.

- In Chapter 5, it is the *users' preferred arrival times* that are continuously distributed.
- In Chapter 6, it is the *users' value of time* that are continuously distributed.

For each game a dedicated method is designed, resulting in a straightforward way to compute the DUE. Each game allows to analyse a specific issue that is of interest in transport science and economic theory. In Chapter 5, the linkage between peaks in demand, understood here as a high density of users preferring to arrive in a small time window and the congestion periods are investigated. For instance we will see that several peaks in demand might merge in a single congested period or alternatively give rise to a congested period each. In Chapter 6, different pricing strategies for one arc of the two-arc network are tested and assessed. A noteworthy result is that under some specific assumptions a profit maximizing toll allows nearly as much welfare gains as a welfare maximizing toll.

## Interest for the rest of the thesis

An important contribution of this Part to the rest of the thesis is the insights it gives on DUE with distributed users' characteristics. Part II demonstrates that correctly representing such an equilibrium is a complex matter and Part IV shows that computing a DUE also is. The experience gained from the analytical resolution of simple examples will be of great help in the design of the computation methods proposed in the last Part.

Chapter 5

# User equilibrium with general distribution of preferred arrival times

The seminal paper on trip scheduling is due to Vickrey (1969), who considered a fixed number of commuters traveling from an origin to a destination by a single route where congestion occurs at a bottleneck, each user being a microeconomic agent minimizing a cost function that involves travel time as well as schedule delay. In the simplest version of the model, Vickrey considered homogeneous users that have same preferred arrival time and same cost function. Many extensions of the model have been provided in the literature, with focus on user heterogeneity. That pertaining to preferred arrival times has been treated by Hendrickson and Kocur (1981) with no solution algorithm. Heterogeneity pertaining to the costs of travel time and of schedule delay has been addressed by e.g. Van der Zijpp and Koolstra (2002), Arnott et al. (1993). Other extensions include the modelling of stochastic demand and capacity, multiple routes or elastic demand - for review see (Arnott et al., 1998).

The known results about the equilibrium pattern of departure times can be summarized as follows. When the preferred arrival time is common to all users, a single congestion period emerges with queue at bottleneck first increasing to a maximum and then vanishing. Smith (1984) and Daganzo (1985) showed that this simple departure pattern holds for a distribution of preferred arrival times, under the so-called "S-shape" assumption of a unique peak period, *i.e.* a single interval on which the density of preferred arrival times exceeds the bottleneck capacity rate. However, in the case of a finite number of preferred arrival schedules and heterogeneous cost functions,

(Lindsey, 2004) and (Van der Zijpp and Koolstra, 2002) showed that the resulting departure pattern may be much more complex with possibly several congestion periods and multiple maxima in queuing time.

The purpose of this chapter is to extend the model of Smith and Daganzo to a general distribution of preferred arrival times. Indeed, this induces a complex pattern of departure times, as in (Van der Zijpp and Koolstra, 2002) and (Lindsey, 2004). The core principle in our analysis is to express the equilibrium distribution of departure times as the solution of a differential equation. This equation involves the distribution of preferred arrival times, as mediated by bottleneck flowing, together with the costs of schedule delay and travel time. The differential equation also inspires a solution algorithm, which consists in searching for the initial instants of queued periods.

The chapter is organized into four main sections and a conclusion. First, Section 1 states the modelling assumptions and provides intuitive reasoning into the structure of the equilibrium pattern. Then, in Section 2 the characteristic differential equation is obtained by mathematical analysis of the optimality conditions. Next, Section 3 states the solution algorithm and provides a theorem of existence of a departure time equilibrium under general distribution of preferred arrival times. Section 4 is devoted to numerical illustration. Lastly some concluding comments are given.

# 1 The model

Consider a single origin-destination pair connected by a single route, and a set of $N$ users with heterogeneous preferred arrival times. In a game-theoretic perspective, every user is modelled as a microeconomic agent seeking unilaterally to minimize his travel cost by adjusting his departure time $h$. This travel cost is parameterized by a travel time function $\tau : h \mapsto \tau(h)$ giving for every instant of entrance the associated travel time on the route. The distribution of individual choices gives rise to a distribution of departure times which makes a cumulated trip volume at the entrance of the route, which may be called the demand. In turn this macroscopic entry trip volume, denoted as $X_+ : h \mapsto X_+(h)$, determines the route travel time $\tau$ on the basis of queuing dynamics. The travel time function $\tau$ represents the supply state. The demand function linking $\tau$ to $X_+$, and the supply function linking $X_+$ to $\tau$, make up a circle of dependency, typical of an equilibrium problem between supply and demand.

This section is purported to specify the assumptions first on the supply side, then on the demand side, so as to state the equilibrium problem in a formal way.

The following notations will be used:

- $\mathcal{H}_+$, $\mathcal{H}_-$ and $\mathcal{H}_p$ respectively are the domains of departure, arrival and preferred times. Without going into the details, let us assume that these are sufficiently large intervals so that no departure nor arrival takes place out of them.

- $X_+$ is a distribution of departure time over $\mathcal{H}_+$ i.e. $X_+$ represents the number of users having departed before $h$ hence also the cumulated trip volume. $X_+$ is assumed to be continuous and differentiable nearly everywhere, with time derivative $x_+(h)$ to be interpreted as the flow rate of departing users at $h$. A last requirement on $X_+$ is that at a maximum instant $h_{Max}$, it holds that $X_+(h_{Max}) = N$ the total number of users.

- $k$ the bottleneck capacity, a flow rate.

- $\tau$ defined on $\mathcal{H}_+$ is a travel time function assumed to be continuous and differentiable nearly everywhere.

- $t$ the function that maps a distribution $X_+$ to a travel time function $\tau$.

- The derivative of function $f$ with respect to the clock time (*i.e.* to a variable $h$) is denoted as $\dot{f}$.

## 1.1 Transport supply - Flowing model

Let us first consider the derivation of travel time function $\tau$ from departure time distribution $X_+$. Travel along the route is assumed unqueued except perhaps at a single bottleneck of deterministic capacity $k$. If the entry flow coming in bottleneck has rate in excess of $k$, then a waiting queue develops where users wait to leave queue according to a First In - First Out (FIFO) discipline. Thus the supply function $t$ is a standard pointwise travel time model. The following relationship in which $Q(h)$ denotes the number of users queuing at $h$ in the bottleneck, and $\tau_0$ is the free flow travel time:

$$t(h) = \tau_0 + \frac{Q(h)}{k} \tag{5.1}$$

where $Q$ stems from the following differential equation:

$$\dot{Q}(h) = \begin{cases} x_+(h) - k & \text{if } Q(h) \neq 0 \text{ or } x_+(h) - k > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

When $X_+$ is continuous, the resulting travel time $\tau$ is well defined and is continuous and differentiable nearly everywhere. Without loss of generality, we assume that $\tau_0 = 0$ thus letting $\tau$ be the queuing time.

The flowing model is represented in a compact way by the following notation:

$$\tau = t[X_+]$$

## 1.2 Demand side

**User behaviour.** Every user is characterized by a preferred arrival time $h_p \in \mathcal{H}_p$ and a travel cost function representing a trade-off between a travel time and a schedule delay, defined as the difference between the actual arrival time $\bar{h}$ and $h_p$. Given travel time function $\tau$, the cost to a user with preferred arrival time $h_p$ upon departing at $h$ is defined as:

$$g(h, h_p; \tau) = \nu\tau(h) + D(h + \tau(h) - h_p) \tag{5.3}$$

where $D$ is the schedule delay cost function and $\nu$ the trade-off between cost and time also referred to as the value of time. Let also assume:

**Assumption I** (On the Schedule Delay Cost function)**.** *The following assumptions are made on $D$.*

*a) $D$ is continuous.*

*b) $D$ is differentiable on $\mathbb{R}$ with derivative $D_l$.*

*c) $D$ is convex.*

*d) $D$ achieves a minimum at $0$ and $D(0) = 0$ .*

These are standard assumptions, (*e.g.* Arnott et al., 1993; Lindsey, 2004) and yield a cost of schedule delay that increases with the lag between actual and preferred arrival time. Assumptions Ic and Id make $D$ to decrease on $\mathbb{R}_-$ and increase on $\mathbb{R}_+$.

Each user is an economic agent modeled as a rational decision-maker with perfect information: he chooses his departure time so as to minimize his cost

function. Given his preferred arrival time $h_p$ and the travel time function $\tau$, his choice of departure time amounts to the following mathematical program:

$$\min_h g(h, h_p; \tau)$$

**The distribution of users.** Consider now a set of $N$ users with a same cost function $g$, but heterogeneous preferred arrival times. This is represented by a cumulative distribution $X_p$ on $\mathcal{H}_p$: $X_p(h_p)$ is the number of those users with preferred arrival time that is less than $h_p$. The derivative of $X_p$, denoted as $x_p$, is defined almost everywhere and is readily interpreted as the flow rate of users with preferred arrival time $h_p$. From its definition, $X_p$ is increasing and semi-continuous. Let also:

**Assumption II** (On the Distribution of Preferred Arrival Times)**.** *The following assumptions are made on $X_p$.*

*a) $X_p$ is continuous.*

*b) $x_p > k$ on a finite number of intervals.*

*c) $x_p \neq k$ almost everywhere.*

Assumption IIb generalizes the "S-shape" assumption considered in Hendrickson and Kocur (1981), Smith (1984) and Daganzo (1985), which could be stated as "$x_p > k$ on a single interval". Those intervals are called *peak periods* as along each of them there are more users that would prefer to arrive than allowed by the route capacity. Intuitively a higher number of peak periods will give rise to a more complex distribution of departure time, with potentially several distinct queuing periods. Assumption IIa is purely technical, so is IIc which is required only to make precise the statement of the algorithms in Section 3.

**The order of departure.** In the literature, little consideration has been given to represent the departure choice decision of a continuous distribution of users. A natural approach is to introduce a departure choice function $H$ mapping a user with preferred arrival time $h_p$ to his chosen departure time $h$. Then distribution $X_+$ stems from:

$$X_+(h) = \int_{\mathcal{H}_p} \mathbf{1}_{H(h_p) \leq h} dX_p(h_p) \tag{5.4}$$

Yet, relation 5.4 is not convenient to handle. For the sake of analytical simplicity, let us assume:

**Assumption III** (On Natural order of departure)**.** *The departure time function is continuous and increasing.*

This implies that users depart in the order of increasing preferred arrival time, and hence is referred to as the natural order assumption. An obvious issue pertains to the existence of an equilibrium choice function which would not satisfy to a natural order. Daganzo (1985) investigated the case with a strictly convex schedule delay costs function and showed that natural order is satisfied by measurable equilibrium choice functions. With barely convex schedule delay costs, all the equilibrium choice functions do not satisfy the natural order, but at least one does (Arnott et al., 1998).

Under the natural order assumption, Equation 5.4 becomes[1]

$$X_+ = X_p \circ H^{-1} \tag{5.5}$$

## 1.3 User Equilibrium statement

Each user tries to minimize his cost function under perfect information. By definition, the user equilibrium (UE) is a situation where no user can reduce his cost by unilaterally changing his decision, here of departure time.

A natural statement of the problem is:

**Definition 5.1** (User equilibrium based on departure time function)**.** *Find an increasing function $H$ such that, letting $X_+ := X_p \circ H^{-1}$:*

$$g(H(h_p), h_p; \tau) \leq g(h', h_p; \tau) \quad \textit{for almost every } h_p \in \mathcal{H}_p,\ h' \in \mathcal{H}_+ \tag{5.6a}$$

$$\tau = t[X_+] \tag{5.6b}$$

The associated distribution of departure times stems from natural order. Equation (5.6a) expresses the impossibility for any user to improve on his departure time decision; Equation (5.6b) is the flowing equation.

Let us provide a simpler, alternative formulation:

[1]For an increasing function $F$ such as $X_+$ or $H$, our definition of its reciprocal function $F^{-1}$ is as follows:

$$F^{-1}(x) := \inf\left\{h : F(h) > x\right\}$$

**Definition 5.2** (User equilibrium based on departure time distribution)**.** *Find an increasing function $X_+$ such that, letting $H_p := X_p^{-1} \circ X_+$:*

$$g(h, H_p(h); \tau) \leq g(h', H_p(h); \tau) \quad \textit{for almost every } h, h' \in \mathcal{H}_+ \tag{5.7a}$$

$$\tau = t[X_+] \tag{5.7b}$$

In (5.7a) the optimality condition is expressed by enumerating the users in order of departure time, whereas in (5.6a) each user is labelled by his preferred arrival time. The relationship between the two arises from the fact that, in natural order, the n-th user to depart is also the n-th user in the order of preferred arrival time. The two problems are equivalent in the following way.

**Proposition 5.3** (Equivalency of equilibrium statements)**.** *(i) A solution $X_+$ of (5.7) yields a solution $H := X_p^{-1} \circ X_+$ of (5.6). (ii) Conversely, if $H$ is a solution of (5.6) then $X_+ := X_p \circ H^{-1}$ is a solution of (5.7).*

*Proof. (i)* Assume that $X_+$ is a solution to (5.7) and consider $H := X_+^{-1} \circ X_p$. Then $H$ is defined, an increasing function of $h$ as the composition of two increasing functions, and $X_+ = X_p \circ H$. Consider $h \in \mathcal{H}_p$ and apply (5.7a) to $h = H(h_p)$: then for all $h' \in \mathcal{H}_+$ it holds that $g(H(h_p), h_p; \tau) \leq g(h', h_p; \tau)$ and hence (5.6a).

*(ii)* Same argument in reverse order. □

This enables us to study the equilibrium by focusing on $X_+$ rather than $H$. In the sequel, we address the UE problem in departure time distribution.

# 2 Properties of the equilibrium departure time distribution

In this section, necessary conditions are derived on an allegedly optimal pattern $X_+$ from the optimality equation (5.7). Then these conditions are shown to be also sufficient. This line of attack had already been taken by Smith (1984), but in the specific case of an S-shape distribution of preferred arrival time.

## 2.1 On queued and peak periods

Assuming that $X_+$ is a solution of the UE problem, let us consider $\tau = t[X_+]$. As $\tau$ is continuous, the sets $\{h : \tau(h) = 0\}$ [resp. $\{h : \tau(h) = 0\}$] are countable unions of closed [resp. open] intervals. We refer to those intervals as *unqueued* [resp. *queued*] periods. Consider first an unqueued period $U$: users departing during $U$ incur only a cost of schedule delay. Thus, it is optimal for a user with preferred arrival time $h_p$ to choose a departure time $h$ interior to $U$ if and only if $h = h_p$. Otherwise he could lower his cost by marginally changing $h$ towards $h_p$. Then at equilibrium $H_p = \mathrm{Id}_{\mathcal{H}_+}$ and $x_+ = x_p$ on $U$.

Now consider a queued period $Q$. As $\tau$ is continuous, non negative and is zero at the endpoints of its definition interval, it has a least one maximum value and possibly minima. The general pattern of travel time is therefore expected to be a sequence of increasing then decreasing sub-periods.

This gives us a crucial insight into the structure of an equilibrium state. First, whenever there is no queue, users arrive (and depart) at their preferred arrival time and thus incur no cost. Second, the peak periods defined above (when $x_p > k$), play an important role in the problem: as unqueued departure flow is equal to scheduled flow at arrival, an unqueued period cannot intersect a peak period except perhaps at isolated points (since $\tau = 0$ cannot be sustained when $x_+ > k$). Therefore, the maximum number of queued periods is bounded by the number of peak periods; whereas the number of unqueued periods is limited to one plus that bound.

To sum up, we have highlighted two important features of $\mathcal{H}_+$ and $\mathcal{H}_p$ under an equilibrium distribution. The set of departure times is divided into alternated periods of unqueued and queued states. Provided that $\mathcal{H}_+$ is "large enough", the first and last periods should be unqueued. To state this principle explicitly, we denote $Q_1 = ]q_0, q_1[, Q_2 = ]q_1, q_2[ \ldots, Q_{2n+1}$ the sequence of unqueued and queued periods, $q_{2k}$ and $q_{2k+1}$ being *transition instants* from an unqueued period to the next queued period, and from queued to unqueued, respectively. Similarly, we denote by $P_1 = ]p_0, p_1[, \ldots, P_{2n+1}$ the sequence of successive peak (when $x_p > k$) and off peak (when $x_p < k$) periods in $\mathcal{H}_p$.

## 2.2 Necessary conditions

Given a solution $X_+$ of the UE problem (5.7), consider the associated functions of travel time $\tau = t[X_+]$, preferred time $H_p = X_+^{-1} \circ X_p$ and cost $g$ (the reference to $\tau$ is omitted for the sake of legibility). Our aim is to turn the optimality conditions on the basis of $g$ into conditions on $X_+$ by means of the flowing equation. To do so, the two states of unqueued versus queued traffic must be addressed as distinct cases.

**About unqueued periods**, we already established that:

$$x_+ = x_p \tag{5.8}$$

and it holds that $\tau(h) = 0$ and $H_p(h) = h$. Then $h = X_+^{-1} \circ X_p(h)$ or equivalently $X_+(h) = X_p(h)$. This applies notably to each instant $q_i$ of transition between queued and unqueued state, yielding that

$$X_+(q_i) = X_p(q_i) \text{ for any } i \in \{0, 1, \dots 2n_q\} \tag{5.9}$$

**About a queued period** $Q$, consider a given $h' \in Q$ together with $H_p(h')$ the preferred arrival time of users departing at $h'$ and let $g^{(h')} : h \mapsto g(h, H_p(h'); \tau)$. As the functions $h \mapsto \tau(h)$ and $h \mapsto D(h + \tau(h) - H_p(h'))$ are differentiable a.e. so is $g^{(h')}$. Denote $\dot{g}^{(h')}(h) = \dfrac{dg}{dh}^{(h')}(h)$. From Equation (5.7a), it must hold $\dot{g}^{(h')}(h) = 0$ for almost every $h' \in \mathcal{H}_+$.

Yet as $D$ is differentiable on $\mathbb{R}^\star$, whenever $h + \tau(h) - H_p(h') \neq 0$:

$$\dot{g}^{(h')} = \nu\dot{\tau}(h) + D_l(h + \tau(h) - H_p(h'))(1 + \dot{\tau}(h')) \tag{5.10}$$

Equation 5.10 is easily extended on $\mathbb{R}$ by defining $D_l(h) := 0$. For $h'$ in $Q$ and $h$ in $\mathcal{H}_p$, we thus have:

$$\nu\dot{\tau}(h) + D_l\big(h + \tau(h) - H_p(h')\big)(1 + \dot{\tau}(h')) = 0 \tag{5.11}$$

Evaluating the previous equation in $h = h'$ and introducing the flowing equation in a queued state, we get that:

$$x_+ = k.\frac{\nu}{D_l(l) + \nu}$$

where $l(h') := h' + \tau(h') - H_p(h')$ is the schedule delay of the user departing at $h$.

Equation (5.11) has two remarkable features. First $x_+ > k$ whenever $l > 0$ and reversely $x_+ > k$ whenever $l > 0$. The function $l$ can be interpreted as the schedule delay incurred by a user departing at $h$. Consequently, each queued period can be divided in early sub-periods when users depart early (that is, depart at a time yielding arrival earlier than preferred ex-ante), during which the entry flow rate is beyond capacity and the queue builds up; and late sub-periods when users depart late, during which the entry flow rate is under capacity and the queue diminishes. Second, (14) can be stated as a differential equation in $X_+$ over $Q_i = ]q_{i-1}; q_i[$. Indeed, according to the flowing equation (5.2) we have $\dot{t} = (x_+ - k)/k$ on $Q_i$, so by integrating over $]q_{i-1}; h[$:

$$\tau(h) + h = q_{i-1} + \frac{X_+(h) - X_+(q_{i-1})}{k}$$

Taking the definition of $H_p = X_p^{-1} \circ X_+$, the lateness $l$ can now be expressed as a function of $X_+$, so that (5.11) yields a differential equation in $X_+$.

To sum up, we have shown that the equilibrium departure time distribution satisfies the differential equations (5.8) and (5.11) respectively on unqueued and queued periods. Successive integrations of these equations along the $Q_i$ periods with an appropriate initial condition coming from the previous period yields the equilibrium departure time distribution, provided that the $Q_i$ periods are given.

## 2.3 Necessary and Sufficient Conditions

Let us now demonstrate that the necessary conditions are also sufficient conditions, owing to the following property:

**Proposition 5.4** (NCS for the UE)**.** *Let $X_+$ be a departure time distribution with associated sequence $Q_i$ of unqueued and queued periods. Then $X_+$ is an equilibrium solution if and only if it satisfies* (5.11) *and* (5.9) *on $Q_{2i}$ and* (5.8) *on $Q_{2i+1}$ for all $i$.*

*Proof of Proposition 5.4.* Having demonstrated the "only if" part in the previous subsection, let us tackle the "if" part by taking a departure time distribution $X_+$ with associated functions $\tau = t[X_+]$ and $H_p = X_p^{-1} \circ X_+$ of travel time and preferred time, respectively.

Assume that $X_+$ satisfies (5.11) and (5.9) on $Q_{2i}$ and (5.9) on $Q_{2i+1}$ for all $i$. Let us fix any $h$ in $\mathcal{H}_+$ and consider the function $g^{(h)} : h' \mapsto$

$\nu\tau(h') + D(h' + \tau(h') - H_p(h))$. The quantity $g^{(h)}(h')$ represents the cost incurred by a user of preferred arrival time $H_p(h)$ when leaving at $h'$. Our aim is to show that $g^{(h)}$ admits a global minimum at $h' = h$. From its definition $g^{(h)}$ is continuous and differentiable almost everywhere, with derivative $\dot{g}^{(h)}$ given by (5.10), with interchange between $h$ and $h'$. Since $H_p$ is an increasing function (as composition of two increasing functions), as is $D_l$ because of the convexity of $D$, $h' \mapsto \dot{g}^{(h)}$ is a decreasing function. Around point $h' = h$ we have that:

$$\dot{g}^{(h)}(h') \gtrless \dot{g}^{(h)}(h) \text{ if } h' \lessgtr h$$

Yet $\dot{g}^{(h)}(h) = 0$ almost everywhere on the basis of either (5.11) in a queued state or (5.8) in an unqueued state. It holds that for almost every $h, h' \in \mathcal{H}_+$,

$$\dot{g}^{(h)}(h') \gtrless 0 \text{ if } h' \gtrless h$$

which means that $h' = h$ is the unique minimum of function $g^{(h)}$. Thus $X_+$ satisfies the optimality condition (5.7a), as well as (5.7b) by assumption. □

## 2.4 Graphical interpretation of the NSC under V-shape schedule delay costs

From here it is assumed that $D$ has the simple, V-shaped form:

$$D(h + t - h_p) = \alpha.(h + t - h_p)^+ + \beta.(h + t - h_p)^- \tag{5.12}$$

where $\alpha$ [resp. $\beta$] are the marginal cost of arriving early [resp. late] with respect to the preferred time $h_p$ and $()^+$ [resp. $()^-$] denotes the positive [resp. negative] part. Under this V-shaped form, equation (5.11) can be restated in the following simple way:

$$x_+(h) = \begin{cases} x_+^E := \dfrac{k\nu}{\nu - \alpha} & \text{if } h + \tau(h) < H_p(h) \\ x_+^L := \dfrac{k\nu}{\nu + \beta} & \text{if } h + \tau(h) > H_p(h) \end{cases} \tag{5.13}$$

Therefore only two departure flows are admissible in a queued period, one made of users planning to arrive early regarding their preferred time and the other of users planning to arrive late. These are denoted by $x_+^E$ and $x_+^L$, respectively, E and L standing for early and late. From their definition $x_+^E > k$ and $x_+^E < k$. Let us now use the cumulated volume representation

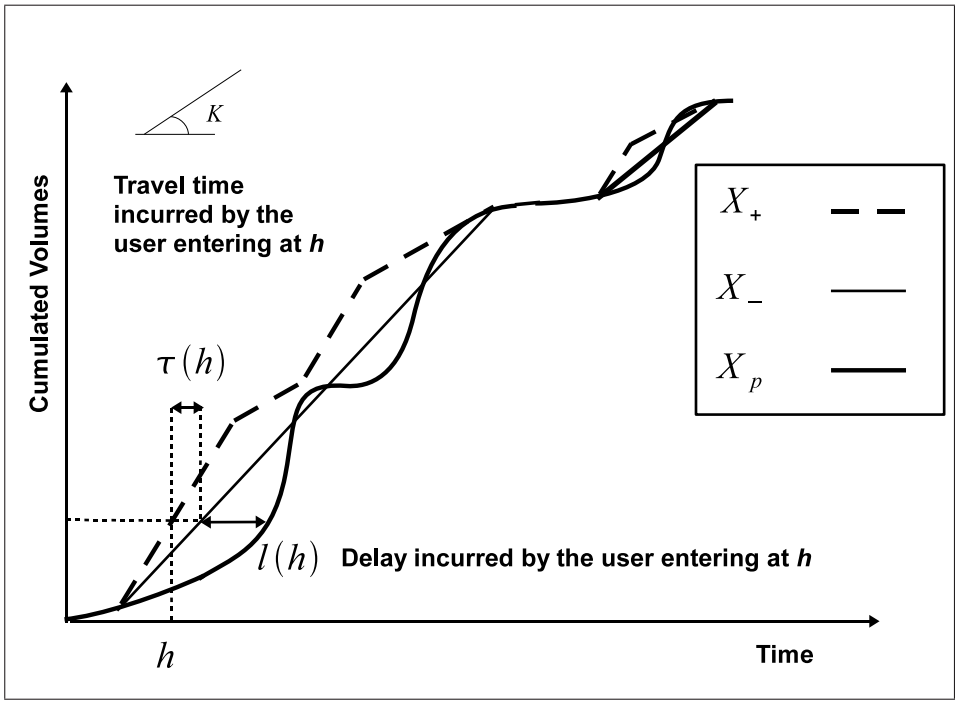



Figure 5.1: Cumulated volume representation of an equilibrium situation

to comment the conditions on $X_+$. Figure 5.1 depicts $X_+$, $H_p$ and $X_- = X_+ \circ (\mathrm{Id}_{\mathcal{H}_+} + t)$, the arrival time distribution.

First, note that $X_-$ can be easily deduced from the sequence of the $Q_i$. Indeed, according to the simple flowing model, the exit flow rate is the capacity $k$ on a queued period and so $X_-$ has slope $k$; out of queued periods $X_-$ simply coincides with $X_p$ and $X_+$. Second, in Figure 5.1 one can read $\tau$ and $l$ from the horizontal distance between respectively the graphs of $X_+$ and $X_-$, and those of $X_-$ and $X_p$. Moreover the intersection points between the graphs of $X_-$ and $X_p$ divide each queued period $Q$ into early and late intervals regarding the preferred arrival time. The transition instants between two successive periods make *critical times at arrival*, denoted as $\bar{h}_i$. Such instants on a period $Q = [q_m; q_M]$ are the solutions of the equation:

$$k.(\bar{h} - h_m) = X_p(\bar{h}) - X_p(h_m) \tag{5.14}$$

Clearly there cannot be more than one $\bar{h}_i$ per peak or off peak period, and their total number over a queued period must be odd. To each critical time at arrival $\bar{h}_i$ let us associate the corresponding departure time $h_i$, so

that they are related by the equation:

$$\bar{h}_i = h_i + \tau(h_i) \tag{5.15}$$

The critical times at departure $h_i$ also divide each queued period $Q_{2i}$ in intervals of earliness or lateness regarding the departure, i.e. in periods where users depart at a time such that they arrive early or late. Those instants correspond to a switch in the departure flow from $x_+^E$ to $x_+^L$ or conversely. Figure 5.2 illustrates the definition of critical times at arrival and at departure.

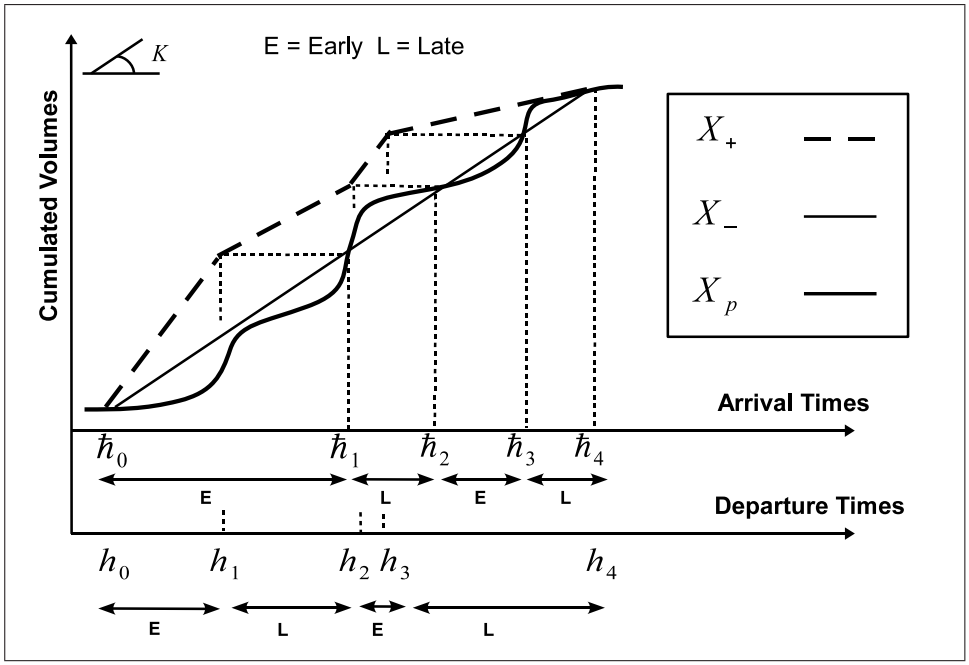



Figure 5.2: Critical times at arrival and at departure

# 3 UE algorithm under V-shaped cost of schedule delay

This section provides an algorithm to compute the equilibrium departure time distribution based on the properties established previously. The objective of the algorithm is to build the distribution of departure time by determining the queued periods. The principle is that, given the beginning of a queued period, both $X_+$ and $\tau$ are easy to compute by integrating equations

(5.11) and (5.1) and stopping when $\tau = 0$: thus the main unknown variable is the initial instant of a queued period, and the algorithm is purported to test candidate initial instants.

Two questions arise about a candidate initial instant. First, will the associated queued period induce an equilibrium state? Then, how to search for all queued periods in such a way as to delimit precisely each of them? Both issues are addressed in an integrated way, by progressive identification of the successive queued periods. A criterion is provided that both guarantees the current queued period to be correct and ensures that the search for the next queued period should focus on later instants. We shall first present an algorithm for testing a candidate initial instant $\hat{\hbar}_0$, then expose the full computation method and next give the proof of convergence. Lastly, based on the algorithm termination we derive the following existence result:

**Theorem 5.5** (Existence of equilibrium)**.** *The user equilibrium problem with general preferred arrival time distribution and V-shaped cost of schedule delay admits at least one solution.*

## 3.1 Testing a candidate initial instant of a queued period

Assuming that a sequence of queued periods has been identified up to time $h_m$, our aim is to identify the initial instant $\hat{\hbar}_0$ of the next queued period, prior to the beginning of the next peak period.

The algorithm is as follows. First equation (5.14) is solved on $[\hat{\hbar}_0; +\infty)$, yielding a sequence of solutions $\hat{\hbar}_i$, which is referred to as the sequence of intersection times at arrival. Then the sequences $(\hat{h}_i)$ and $(\hat{t}_i)$ are derived in a recursive way, by setting initial value to $\hat{h}_0 = \hat{\hbar}_0$ and $\hat{t}_0 = 0$, and then by using the following, recursive formulae:

$$x_+^i.(\hat{\hbar}_{i+1} - \hat{\hbar}_i) := k.(\hat{\hbar}_{i+1} - \hat{\hbar}_i) \tag{5.16}$$

$$\text{with } x_+^i = x_+^E \text{ if } i \text{ is even and } x_+^L \text{ otherwise,}$$

and,

$$\hat{t}_i := \hat{\hbar}_i - \hat{h}_i \tag{5.17}$$

The sequences $(\hat{\hbar}_i)$, $(\hat{h}_i)$ and $(\hat{t}_i)$ are purely geometric constructions, as illustrated in Figure 5.3. Yet intuitively $(\hat{\hbar}_i)$ and $(\hat{h}_i)$ would correspond to

the i-th critical times at arrival and departure derived from a given candidate $\hat{\hbar}_0$ and $(\hat{t}_i)$ to the corresponding travel times. They define a candidate distribution $\hat{X}_+$ that a priori is not flow-consistent with the candidate arrival time distribution $\hat{X}_-$. Two unphysical phenomena may occur:

- "Travel time becomes negative": for some $i$, $\hat{\hbar}_i < \hat{h}_i$ or equivalently $\hat{t}_i < 0$. This typically corresponds to a situation where the candidate queued period started too early.
- "Queue does not vanish": for all $i$, $\hat{\hbar}_i > \hat{h}_i$ or equivalently $\hat{t}_i > 0$, which corresponds to a situation where the candidate queued period started too late.

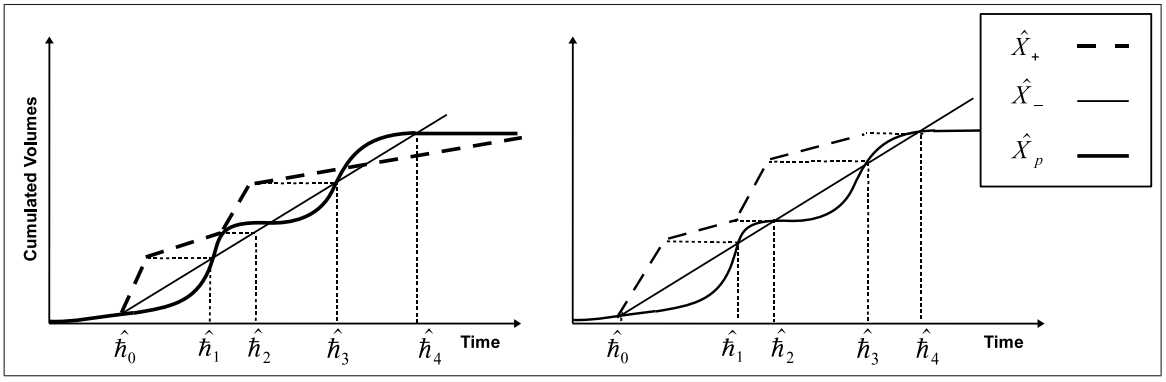



Figure 5.3: Testing a candidate initial instant

We claim that the sequence $(\hat{t}_i)$ allows us to assess the suitability of $\hat{\hbar}_0$ as initial instant of queuing in an equilibrium state. The intuition is as follows: assume that there exists $j$ such that $\hat{t}_j = 0$ and $\hat{\tau}_i > 0$ for $i < j$. Then, by deriving $X_+$ from the sequence $(\hat{h}_i)_{i \leq j}$, (5.11) hold on $Q = [q, \hat{h}_i]$ and $Q$ indeed describes a queued period. Therefore, the condition "$\exists j$ such as $\hat{\tau}_j = 0$ and $\hat{\tau}_i \geq 0$ for $i < j$ " is a necessary condition for $\hat{\hbar}_0$ to be the instant we are looking for. Yet, it will be seen later on to be too weak for sufficiency; the appropriate criterion is in fact "$\exists j$ such as $\hat{\tau}_j = 0$ and $\hat{\tau}_i \geq 0$ for all $i$" or equivalently "$\min_i \hat{\tau}_i = 0$ ". Intuitively, this guarantees that the candidate queued period "leaves enough space" for the subsequent ones. The algorithm is stated below in explicit pseudo-code.

**Algorithm 5.1** `QTest`$(\hat{\hbar}_0)$

**Inputs:** A candidate initial instant $\hat{\hbar}_0$
**Outputs:** $h_e$, $\min_i \bar{t}_i$
**Set** $\bar{t}_0$ to 0 and $\hat{h}_0$ to $\hat{\hbar}_0$ and **Set** the $n$ solutions to the sequence $(\hat{\hbar}_i)_{i=0\ldots n-1}$ in increasing order.
**For** $i = 1 \ldots n-1$ **do**
**Set** $\hat{\hbar}_i := \hat{\hbar}_{i-1} + k/x^i_+.(\hat{\hbar}_i - \hat{\hbar}_{i-1})$
**Set** $\hat{\tau}_i :=$
**End For**
**Set** $k$ to $\arg\min_i \hat{\tau}_i$ and **Set** $h_e$ to $\hat{h}_k$

## 3.2 Main algorithm

The general philosophy of our method is to find successively the queued periods in the UE departure time distribution, starting from the first peak period. Algorithm 5.2 consists in searching over an interval $[h_m, h_M]$ for the initial instant of a queued period, by testing candidate initial instants $\hat{\hbar}_0$ on the basis of Algorithm 5.1. The search method is a dichotomy process oriented by the sign of $\min_i \hat{\tau}_i = 0$. Algorithm 35.3 uses Algorithm 5.2 repeatedly until all peak periods have been addressed; it returns the sequence of queued periods which fully determines $X_+$. The computation process is illustrated in Figure 5.4.

**Algorithm 5.2** `findqueuedPeriod`$([h_m, h_M])$

**Inputs:** A search period $[h_m, h_M]$
**Outputs:** $[q_m, q_M]$
**Parameters** $\epsilon$ a tolerance level
**Repeat**
**Set** $q_m := (h_m + h_M)/2$
**Set** $\{q_e, \min \tau\}$ to `QTest`$(q_m)$
**If** $\min \tau > 0$ **then** Set $h_M := q_m$
**else Set** $h_m := q_m$
**Until** $|\min \tau| < \epsilon$

**Algorithm 5.3** `equilibriumComputation`($\mathcal{H}_+$)

**Inputs:** The set of admissible departure time $\mathcal{H}_+$
**Outputs:** The sequence of queued periods $Q_{2k}$
**Set** $k := 1$
**Set** $h_M$ to the initial instant of the first period
**Set** $h_m := \inf \mathcal{H}_+$
**Repeat**
**Set** $Q_{2k}$ to `findqueuedPeriod`($[h_m; h_M]$)
**Set** $k := k + 1$
**Set** $h_m := \sup Q_{2k}$
**Set** $h_M$ to the initial instant of the first period after $Q_{2k}$
**Until** there is no peak after $h_m$

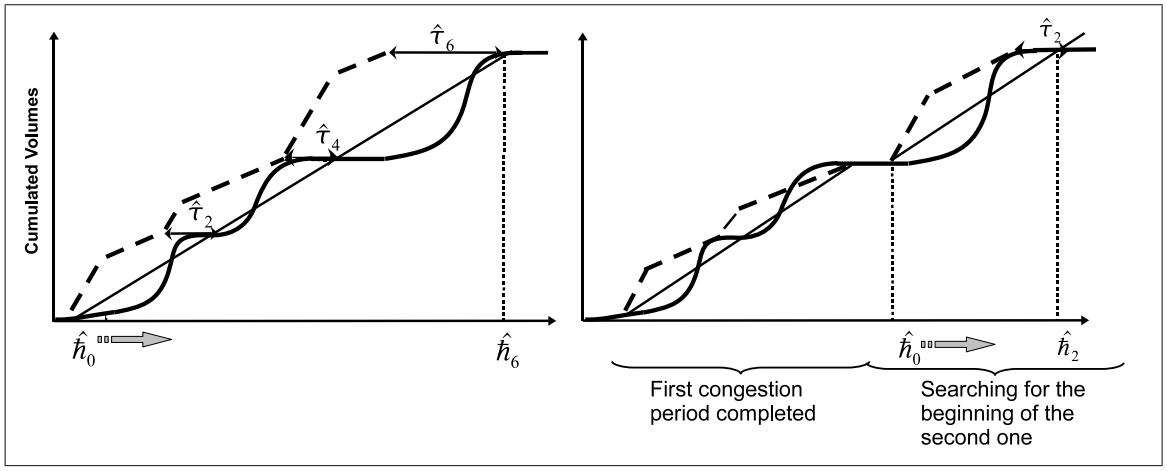



Figure 5.4: User equilibrium algorithm

## 3.3 Proofs

Consider the functions $\hat{\tau}_i(h_0)$ defined by (5.16) and (5.17) on a given period $[h_m, h_M]$. The proofs of existence and termination essentially derive from the following property.

**Proposition 5.6.** *$W_m(h_0) := \min_i \hat{\tau}_i(h_0)$ is a continuous and decreasing function.*

The proof of Proposition 5.6 is given in Appendix A.

The proposition implies that the equation $W_m = 0$ has a solution on $[h_m, h_M]$ if "$W_m(h_m) \geq 0$ and $W_m(h_M) \leq 0$". Then Algorithm 5.2 applied

to an off-peak period with adequate inputs must terminate and yield a suitable initial crtitical instant $h_0$. Moreover, by progressive identification of the successive queued periods in the equilibrium state, Algorithm 5.3 must terminate.

Let us finally address the issue of existence for an equilibrium departure time distribution.

*Proof of Theorem 5.5.* Consider the departure time distribution $X_+$ computed from the outputs $(Q_{2k})$ of Algorithm 5.3 together with its associated functions $\tau$ and $H_p$ of travel time and preferred time, respectively. Then for all $k$, $\tau \geq 0$ on $Q_{2k}$ and $\tau = 0$ elsewhere. Moreover $X_+$ satisfies (5.11) and (5.9) by construction on queued periods and (5.8) on unqueued periods. The existence theorem then follows directly from Proposition 5.4. □

# 4 Numerical experiments

Having implemented the algorithm in a computer program under the Scilab environment (Scilab Consortium, 2010), a series of numerical experiments were performed by progressively moving two peak periods closer to each other (Figure 5.5). Initially there are two distinct queued periods, each of them with a single maximum of travel time. Then the two periods are merged into a single one with two maxima. Further, when the peak periods are close enough, the two maxima collapse into a single one yielding the same pattern as with a single peak period: the well-known pattern made up of one queue-loading sub-period followed by an unloading one.

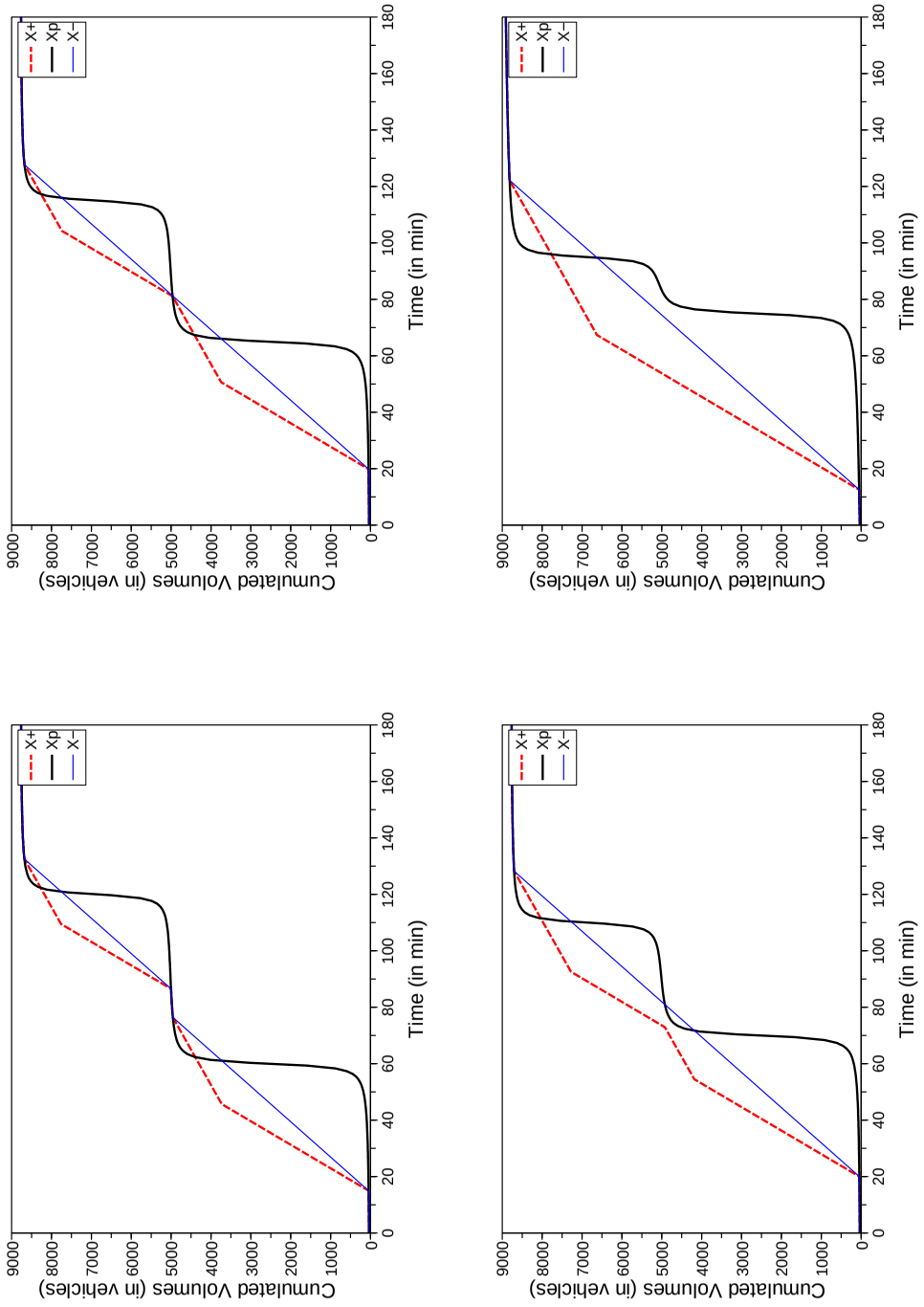



Figure 5.5: Numerical experiments

# 5 Conclusion

This chapter showed that relaxing the S-shape assumption on the pattern of preferred arrival times in the single bottleneck may give rise to a much more complex pattern of departure times, with potentially several queued periods and travel time maxima. Applications of such a model may include the assessment of transportation policies, such as congestion pricing or flextime promotion.

Among the improvements that would make sense, a major one is to introduce heterogeneity in the cost of schedule delay. Indeed, complex road pricing schemes are based on the principle that one can segregate high schedule costs from lower ones by imposing time varying tolls. Therefore the heterogeneity in schedule delay cost functions and in the user cost of time is essential in assessing the benefits of such schemes.

Chapter 6

# User equilibrium with continuously distributed values of time

Facing ever rising levels of traffic congestion, many local authorities have given serious consideration to road pricing. Among the existing schemes, value pricing has enjoyed a reasonable success especially in the US. A famous instance is the one currently operating on the SR91 in California. In a value pricing scheme, travellers choose between two roadways: one is free but congested, while the other one is priced but free flowing.

An important literature has already explored the design and the assessment of value pricing. It has mainly focused on two features of the problem. On one hand, it has been pointed out that the welfare gain is highly affected by the level of heterogeneity among the travellers, especially regarding their value of time (*e.g.* Papon, 1992; Verhoef and Small, 1999; Small and Yan, 2001). These results have been achieved using static models, thus neglecting the time-varying nature of congestion and the possibility for travellers to adjust their departure time, *e.g.* by leaving earlier than preferred to avoid traffic jams. On the other hand, dynamic models of congestion, most of them inspired by Vickrey's bottleneck model, have been used to assess value pricing. Most of these works tend to neglect the heterogeneity among travellers, or to have a crude representation of it, for instance by considering only two possible values of times (*e.g.* De Palma and Lindsey, 2002). Papers accounting for both aspects are very rare.

A notable exception is van den Berg and Verhoef (2010) who considered a bottleneck model with two routes, where heterogeneity is represented by a continuum of values of time. Under this framework, they assessed two pricing

policies. In both of them a time-varying and queue clearing toll is set on one roadway, while the other one is untolled. However in the first policy the toll is set to maximize revenue while in the second it is set to maximize the welfare gains. This gives very interesting results: as in the static case, the level of heterogeneity impacts the welfare gains of both schemes but in an adverse way. In the static case the relative efficiency of a private pricing scheme decreases with heterogeneity. With a bottleneck model, it is the contrary. This is saying that changing the representation of congestion from a static to a dynamic framework leads to radically different results when it comes to the impact of heterogeneity.

Now Van den Berg and Verhoef's treatment of the problem is only valid for time-varying and queue clearing tolls. This chapter aims at investigating in an analytical approach if their result stands for flat tolls. Indeed, in practice, a fully time-varying toll is rarely set and tolling schemes usually have one fixed toll during the entire peak or day. To achieve this goal, it is required to state and derive a dynamic user equilibrium model for two route networks, which leads to some reasonably complex analytic.

The chapter is structured in three sections and a conclusion. In the first section, a bottleneck model with one route and continuous heterogeneity in the value of time is exposed. It is shown that the problem can be reduced to the resolution of a differential equation and examples of resolution are given in the case of a uniform distribution of the values of time. The second section extends the model for two routes. Finally, the results are exploited to assess a value pricing scheme under two ownership regimes (Section 3). In the first case the priced roadway is publicly owned and the toll is set to maximise the welfare gains; in the second one, it is privately owned and the toll is set to maximise profits.

# 1 Model with one route

## 1.1 Model statement

Consider a single OD pair connected by a single route of deterministic capacity $k$. A set of users wish to go from the origin to the destination and prefer to arrive at a given instant $h$. All the users share the same preferred arrival instant $h_p$ but they might proceed to a trade-off between effectively arriving at that instant and avoiding high travel times. This trade-off depends on

how much they value a decrease in travel time and a reduction in schedule delay, the latter being defined as the lag between their effective arrival time and their preferred one.

For a given arrival pattern, the congestion on the route is modelled by a travel time function $\tau : \mathbb{R} \to \mathbb{R}_+$ that associates to an arrival time $h$, the travel time $\tau(h)$ required to arrive at that time.

The following alineas describe how the users and the equilibrium are represented.

**Demand** Users only differ w.r.t. their value of (travel) time (VoT). They have the same preferred arrival time and schedule delay cost. Users are modelled as a continuum: $V$ is a cumulative distribution representing a set of users whose value of time varies within $[\nu_m, \nu_M]$. The quantity $V(\nu)$ is the volume of users whose value of time is under $\nu$. The distribution $V$ is assumed continuously differentiable and its derivative is denoted $v$. Given a travel time $\tau$, a user with a value of time $\nu$ within $[\nu_m, \nu_M]$ appraises the option of *arriving* at $h$ using the following cost function:

$$g(h; \nu, \tau) = \nu\tau(h) + \alpha(h - h_p)^- + \beta(h - h_p)^+ \tag{6.1}$$

where $(.)^-$ and $(.)^+$ denote the negative and the positive part. Equation (6.1) expresses a trade off between the travel time and the arithmetical lateness with respect to a preferred arrival time.

In our framework *an assignment of the demand* is described by a pair of functions $(h_-, h_+)$, where $h_- : [\nu_m, \nu_M] \to ]-\infty, h_p]$ and is increasing, and $h_+ : [\nu_m, \nu_M] \to [h_p, +\infty[$ and is decreasing. That is to say that for each value of time $\nu$ in $[\nu_m, \nu_M]$, users divide themselves in two categories: part of them will decide to arrive before $h_p$ while the others will arrive after.

Without loss of generality, $h_p$ is set to 0 for the rest of the chapter.

**Supply** Travel time on the route is assumed to follow standard pointwise bottleneck model with zero free flow travel time (see Chapter 4). Recall that a bottleneck model can be compactly represented by a function $t$ that maps a cumulated inflow $X_+$ in the bottleneck to a travel time function $\tau = t[X_+]$.

We say that a function $\tau$ is a $V$-feasible travel time if there exists a cumulated flow $X_+$ such that (1) $X_+(+\infty) = V(\nu_m)$ and that (2) $t[X_+] = \tau$. As a consequence of the properties of the bottleneck model, all $V$-feasible travel time $\tau$ are continuous, differentiable nearly everywhere and $\dot{\tau} \geq -1$.

Moreover, $\tau$ has a compact support *i.e.* there exists a bounded interval $I$ such that $h \notin I$ implies $\tau(h) = 0$.

**Dynamic User Equilibrium problem** Solving the general equilibrium problem would imply to consider a distribution on the space $[\nu_m, \nu_M] \times \mathcal{H}$. In this chapter, we will focus on a special type of user equilibrium where for a given value of time, the corresponding users are assigned to two arrival times, one before $h_p$ and the other after. We state the equilibrium formally as follows:

**Definition 6.1** (Dynamic User Equilibrium problem (DUE))**.** *Find a state of the demand $(h_-, h_+)$, and a $V$-feasible travel time $\tau$ whose support is $[h_-(\nu_M), h_+(\nu_M)]$, such that:*

$$g\big(h_-(\nu);\nu,\tau\big) = \min_h \; g(h;\nu,\tau) \quad \forall \nu \tag{6.2a}$$

$$g\big(h_+(\nu);\nu,\tau\big) = \min_h \; g(h;\nu,\tau) \quad \forall \nu \tag{6.2b}$$

$$k.\big(h_+(\nu) - h_-(\nu)\big) = V(\nu) \quad \forall \nu \tag{6.2c}$$

The interpretation is as follows. Equations (6.2a) and (6.2b) express the optimality of the solution while Equation (6.2c) is the constraint imposed by the bounded capacity at exit. For the sake of analytical simplicity, we will look for solutions $(h_-, h_+, \tau)$ of the DUE that are continuously differentiable almost everywhere.

## 1.2 Comments about the equilibrium formulation

**Scope of the formulation** The DUE problem, as presented above, states *a* specific type of equilibrium and has an implicit assumption embedded in its definition. Users arrive according to a specific discipline: a user with VoT $\nu$ has only two optimal arrival instants, one before $h_p = 0$, given by $h_-(\nu)$ and one after, given by $h_+(\nu)$ (see Equations (6.2a) and (6.2b)). Moreover users arrive in the order of their VoT before $h_p$ and in the reverse order after $h_p$ (as $h_-$ and $h_+$ are respectively increasing and decreasing by definition of an assignment of the demand). This implies an equilibrium structure with a *single peak in travel time*, centred on $h_p$ and where *users with high VoT arrive near the preferred arrival time* while users with lower VoT arrive on the flanks of the peak period.

A formal approach to justify this assumption would be to formulate the equilibrium in a larger framework, as the dynamic congestion games introduced in Chapter 6 and show the assumption holds. Now the following two-step qualitative reasoning shows that the equilibrium structure assumed here is sensible:

1. *In equilibriums with a single peak in travel time centred in $h_p$, the assumption hold.* Indeed, with this structure of travel times, a user with high VoT has always more incentive to arrive closer to $h_p$, where the the travel times are higher, than a one with low VoT. The order of arrival described by a pair of functions $(h_-, h_+)$ is the only compatible with an equilibrium.

2. *No equilibrium with multiple peaks in travel times exists.* If multiple peaks in travel times exist, there at least one peak at an instant $h \neq h_p$. Assume there exists an equilibrium with a peak at $h < h_p$. Then the user arriving at $h$ has an incentive to arrive slightly later, at an instant $h + \delta h$ closer to $h_p$ and at which the travel time is lower. This travel time structure is incompatible with an equilibrium. The case $h > h_p$ is similar.

**From an arrival time to a departure time perspective** The DUE's formulation retained here is based on the arrival time functions $h_+$ and $h_-$. On the contrary the one presented in the previous chapter was based on departure time functions. This latter point of view is more intuitive: from a behavioural perspective it is simpler to consider that users choose their departure time and that their arrival time results from the FIFO queue at the bottleneck. On the contrary, when considering the arrival time as users' choice variable, one has to impose Equation (6.2c) to guarantee the physical constraint of the bottleneck on the outcoming flow. However the arrival time approach leads to simple analytics and that's why it was chosen here.

## 1.3 Derivation

The philosophy of our derivation method is the following. An additional quantity, the map $\nu \to \tilde{g}(\nu)$, is first introduced. $\tilde{g}(\nu)$ physical interpretation is the generalized cost incurred by a user with value of time $\nu$. It is then

showed that the DUE problem is equivalent to the second order differential equation in $\tilde{g}$ presented in Definition 6.2.

**Definition 6.2** (Equilibrium Cost Problem (ECP))**.** *Solve the following differential equation on* $[\nu_m, \nu_M]$*:*

$$\frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu) = -\frac{\alpha\beta}{k.(\alpha+\beta)}.\frac{v(\nu)}{\nu} \tag{6.3}$$

*with boundary conditions:*

$$\frac{\partial \tilde{g}}{\partial \nu}(\nu_M) = 0 \tag{6.4}$$

$$\tilde{g}(\nu_M) = \frac{\alpha\beta}{k.(\alpha+\beta)}.V(\nu_M) \tag{6.5}$$

The Equilibrium Cost Problem (ECP) is a simple second order differential equation that admits a unique solution. We claim that the ECP is equivalent to the DUE problem in the following sense.

**Proposition 6.3** (Equivalency of the DUE and the ECP)**.**

*(i) If* $(h_-, h_+, \tau)$ *solves the DUE problem, then* $\tilde{g}(\nu) := \nu.\tau(h_-(\nu)) - \alpha.h_-(\nu)$*, or equivalently* $\tilde{g}(\nu) := \nu.\tau(h_+(\nu)) + \beta.h_+(\nu)$*, solves the ECP.*

*(ii) If* $\tilde{g}$ *solves the ECP then the triple* $(h_-, h_+, \tau)$ *defined by:*

$$h_-(\nu) := -\frac{\alpha}{k.(\alpha+\beta)}V(\nu) \tag{6.6}$$

$$h_+(\nu) := \frac{\beta}{k.(\alpha+\beta)}V(\nu) \tag{6.7}$$

$$\tau(h) := \frac{\partial \tilde{g}}{\partial \nu}\big(h_-^{-1}(h)\big) \tag{6.8}$$

*is a solution to the DUE problem.*

## 1.4 Proof of the equivalency result

The two following alineas give the proof of Proposition 6.3. The first one deals with the part $(i)$ of the proposition while the second one deals with the part $(ii)$. Although the proof can be omitted without loss of continuity, it is not devoid of interest and it gives useful insights into the equilibrium structure.

**Necessary conditions for the equilibrium** Assume given $(h_-, h_+, \tau)$ a solution to the equilibrium travel time problem such that $h_-, h_+$ and $\tau$ are continuously differentiable. First we define:

$$\tilde{g}(\nu) := \nu.\tau(h_-(\nu)) - \alpha.h_-(\nu) = \nu.\tau(h_+(\nu)) + \beta.h_+(\nu) \tag{6.9}$$

$\tilde{g}(\nu)$ thus gives the cost incurred by a user with value of time $\nu$. Note that Equation (6.9) is well defined as $(h_-, h_+, \tau)$ is a solution of (6.2a) and (6.2b).

Then remark that $g(.;\nu,\tau)$ has the following property as a consequence of Equations (6.2a) and (6.2b):

$$\frac{\partial g}{\partial h}(h_+(\nu);\nu,\tau) = \frac{\partial g}{\partial h}(h_-(\nu);\nu,\tau) = 0$$

Consequently:

$$\begin{cases} \dfrac{\partial \tau}{\partial h}(h_-(\nu)) &= \dfrac{\alpha}{\nu} \\ \dfrac{\partial \tau}{\partial h}(h_+(\nu)) &= -\dfrac{\beta}{\nu} \end{cases} \tag{6.10}$$

Equations (6.10) and (6.2c) can be used to derive an equation on $\tilde{g}(\nu)$. Indeed:

$$\begin{cases} \dfrac{\partial \tilde{g}}{\partial \nu}(\nu) = \tau(h_-(\nu)) \;\Rightarrow\; \dfrac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu) = \dfrac{\partial \tau}{\partial h}(h_-(\nu))\dfrac{\partial h_-}{\partial \nu}(\nu) \\ \dfrac{\partial \tilde{g}}{\partial \nu}(\nu) = \tau(h_+(\nu)) \;\Rightarrow\; \dfrac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu) = \dfrac{\partial \tau}{\partial h}(h_+(\nu))\dfrac{\partial h_+}{\partial \nu}(\nu) \end{cases} \tag{6.11}$$

Combining the Equations (6.11) and (6.10) then yields:

$$\frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu) = \frac{\alpha\beta}{\nu(\alpha+\beta)}\left(\frac{\partial h_-}{\partial \nu} - \frac{\partial h_+}{\partial \nu}\right)$$

Using (6.2c) we finally get the equation of Definition 6.2:

$$\frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu) = -\frac{\alpha\beta}{k.(\alpha+\beta)}.\frac{v(\nu)}{\nu}$$

This is a simple second order differential equation that can be easily solved knowing the two boundary conditions:

$$\frac{\partial \tilde{g}}{\partial \nu}(\nu_M) = 0$$

$$\tilde{g}(\nu_M) = \frac{\alpha\beta}{k.(\alpha+\beta)}.V(\nu_M)$$

The top equation comes from the fact that $\tau\big(h_+(\nu_M)\big) = 0$ and Equation (6.11). The bottom equation comes from the Equations (6.2c) and (6.9) applied in $\nu_M$.

Finally note that the quantities $h_+$, $h_-$ and $\tau$ can be expressed directly with respect to $\tilde{g}$. Combining Equations (6.11) and (6.10) yields:

$$\begin{cases} \dfrac{\partial h_-}{\partial \nu}(\nu) = -\dfrac{\alpha}{k.(\alpha+\beta)}v(\nu) \\ \\ \dfrac{\partial h_+}{\partial \nu}(\nu) = \dfrac{\beta}{k.(\alpha+\beta)}v(\nu) \end{cases}$$

Note that $\tau(h_-(\nu_M)) = \tau(h_-(\nu_M)) = 0$ as required in the DUE definition for an equilibrium travel time function, so $h_-(\nu_M) = -\tilde{g}(\nu_M)/\alpha$ and $h_+(\nu_M) = \tilde{g}(\nu_M)/\beta$. Integrating the previous equations with these boundary conditions yields:

$$\begin{cases} h_-(\nu) = -\dfrac{\tilde{g}(\nu_M)}{\alpha} + \dfrac{\beta}{k.(\alpha+\beta)}(V(\nu_M) - V(\nu)) = -\dfrac{\beta}{k.(\alpha+\beta)}V(\nu) \\ \\ h_+(\nu) = \dfrac{\tilde{g}(\nu_M)}{\beta} + \dfrac{\alpha}{k.(\alpha+\beta)}(V(\nu) - V(\nu_M)) = \dfrac{\alpha}{k.(\alpha+\beta)}V(\nu) \end{cases}$$

Finally, from Equation (6.10):

$$\tau(h) = \frac{\partial \tilde{g}}{\partial \nu}\big(h_-^{-1}(h)\big) = -\frac{\partial \tilde{g}}{\partial \nu}\big(h_+^{-1}(h)\big)$$

**Sufficiency conditions for the equilibrium** Consider $\tilde{g}$ the solution to the ECP and let $h_-$, $h_+$ and $\tau$ be defined by Equations (6.6), (6.7) and (6.8). They have the following properties: $\tau$ is defined on $[h_-(\nu_M), h_+(\nu_M)]$, and is continuous on this interval and differentiable on $[h_-(\nu_M), 0[$ and $]0, h_+(\nu_M)]$. The functions $(h_-, h_+)$ are continuous and differentiable on $[\nu_m, \nu_M]$. Moreover $\tau(h_-(\nu_M)) = \tau(h_+(\nu_M)) = 0$.

Let us prove that the triple $(h_-, h_+, \tau)$ thus defined is a solution to the DUE problem. The proof proceeds by demonstrating the three following claims for a given $\nu$.

**Claim 1** $g(h_-(\nu); \nu, \tau) = g(h_+(\nu); \tau, \nu)$

This is straightforward by definition of $(h_-, h_+)$ from Equation (1.4) and by definition of $g(.;\nu,\tau)$.

**Claim 2** $\dfrac{\partial g}{\partial h}(h_-(\nu);\nu,\tau) = \dfrac{\partial g}{\partial h}(h_+(\nu);\nu,\tau) = 0$

Note that

$$\frac{\partial \tilde{g}}{\partial \nu}(\nu) = \tau(h_-(\nu)) \Rightarrow \frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu) = \frac{\partial \tau}{\partial h}(h_-(\nu))\frac{\partial h_-}{\partial \nu}(\nu)$$

The quantity $\tilde{g}$ is a solution of the ECP and thus satisfies (6.3). $h_-$ is defined by (6.6), so we have:

$$\frac{\partial \tau}{\partial h}(h_-(\nu)) = -\frac{\alpha}{\nu}$$

Finally:

$$\frac{\partial g}{\partial h}(h_-(\nu);\nu,\tau) = -\alpha + \nu\frac{\partial \tau}{\partial h}(h) = 0$$

Using the same arguments, it can be shown that $\dfrac{\partial g}{\partial h}(h_+(\nu);\nu,\tau) = 0$.

**Claim 3** $g(h_-(\nu);\nu,\tau) = \min\limits_{h\in\mathbb{R}^\star_-} g(h;\nu,\tau)$ **and** $g(h_+(\nu);\nu,\tau) = \min\limits_{h\in\mathbb{R}^\star_+} g(h;\nu,\tau)$

As the pair $(h_-, h_+)$ satisfies Equations (6.10), they are respectively decreasing and increasing functions. From Equations (6.11), it comes that $\partial\tau/\partial h$ is increasing on $\mathbb{R}^\star_-$ and $\mathbb{R}^\star_+$ and thus that $\tau$ is convex on these two intervals. $g(.;\nu,\tau)$ is hence clearly convex on $\mathbb{R}^\star_-$ and $\mathbb{R}^\star_+$. As $h_-(\nu)$ and $h_+(\nu)$ are local minimums of $g(.;\nu,\tau)$ (Claim 2), it yields that they are global minimums, respectively on $\mathbb{R}^\star_-$ and $\mathbb{R}^\star_+$.

From Claims 1, 2 and 3, the triple $(h_-, h_+, \tau)$ satisfies Equations (6.2a) and (6.2b). Equation (6.2c) is straightforward.

## 1.5 General properties of the DUE

The equivalency result of Proposition 6.3 directly leads to some general properties of the user equilibrium. Some of them are listed below.

**Property 6.4** (Existence and uniqueness of the DUE)**.** *There exists a unique DUE as expressed in Definition 6.1.*

This is a direct consequence of Proposition 6.3 and of the existence and uniqueness of the solutions of the ECP.

**Property 6.5** (On the equilibrium cost properties)**.** *The function $\tilde{g}$ is convex and increasing.*

According to Property 6.5, users with the highest VoT incur the highest costs. As on the contrary they incur the lower travel time costs, this implies that schedule delay costs decrease with the VoT at a lower rate than travel time costs.

*Proof of Property 6.5.* The function $\tilde{g}$ is convex as a solution of the ECP. It is increasing as $\partial\tilde{g}/\partial\nu$ is increasing and $\partial\tilde{g}/\partial\nu(\nu_M) = 0$. □

**Property 6.6** (On the sensitivity to the distribution of VoT)**.** *The two following quantities depend on $V(\nu_M)$ but not directly on $V$:*

- *the cost incurred by the users of maximum VoT, i.e.* $\tilde{g}(\nu_M) = \frac{\alpha\beta}{k.(\alpha+\beta)}.V(\nu_M)$,
- *the support of* $\tau$*, i.e.* $[h_-(\nu_M), h_+(\nu_M)] = \left[-\frac{\alpha}{k.(\alpha+\beta)}V(\nu_M), \frac{\beta}{k.(\alpha+\beta)}V(\nu_M)\right]$.

The very object of this chapter is to investigate the impact of user heterogeneity on the equilibrium properties. In this context, Property 6.6 is especially interesting and rather surprising. Naturally the expression of $\tilde{g}(\nu_M)$ is the generalized cost in Vickrey's original model with homogeneous users. Note that $\tilde{g}(\nu_M)$ is also the maximum value incurred by any user.

## 1.6 Application in the case of a uniform distribution

In this subsection we set $V(\nu) = \theta.(\nu - \nu_m)$, hence choosing a uniform repartition of the values of time. Equation (6.3) then becomes:

$$\frac{\partial^2\tilde{g}}{\partial\nu^2} = -\frac{\alpha\beta}{k(\alpha+\beta)}.\frac{\theta}{\nu}$$

The integration is straightforward:

$$\tilde{g}(\nu) = -\theta\eta\nu\ln(\nu/\nu_M) + \theta\eta(\nu_M - \nu) + \tilde{g}(\nu_M)$$

letting $\eta = \alpha\beta/k(\alpha+\beta)$. $\tilde{g}$ is an increasing function of $\nu$, which implies that at equilibrium, users with high values of times incur the highest generalized costs. It is not that straightforward as, on the contrary, they experience a lower travel time. Indeed from Equations (6.11) and (1.6) we get:

$$\tau(h_-(\nu)) = \tau(h_+(\nu)) = -\theta\eta\ln(\nu/\nu_M) \tag{6.12}$$

which is clearly decreasing with $\nu$. Note that these two remarks are still true in the general case. At equilibrium users with high value of time experience higher costs but lower travel times.

The two figures below illustrate the influence of the heterogeneity w.r.t value of time. The numerical setting is the following: we considered a population of 9000 users undertaking a route with a bounded capacity of 3600 pcu/hour. This means that the peak lasts 2.5 hours. The average value of time is 8 euros and it is spread uniformly across the population between $\nu_m$ and $\nu_M$. The scheduling cost parameters are $\alpha = 4$ and $\beta = 15.6$. This values have been chosen to get a comparable framework as in van den Berg and Verhoef (2010).

Figure 6.1 and Figure 6.2 show respectively the generalized costs $\tilde{g}$ as a function of the value of time and the travel time as a function of the entrance time for an equilibrium situation. Several distributions are tested, each of them with a different spread (*i.e.* with a different value for $(\nu_M - \nu_m)$) but sharing the same average value. This allows seeing the sensitivity of the equilibrium to user heterogeneity.

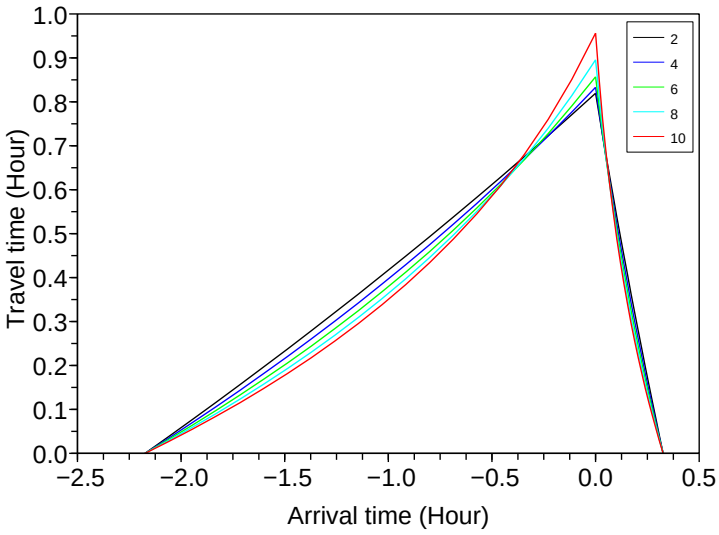



Figure 6.1: Travel time for different spreads of the values of time as a function of arrival time

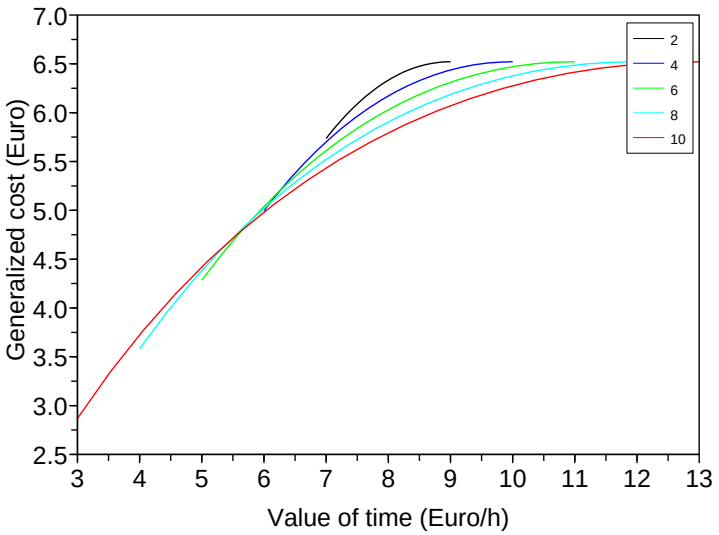



Figure 6.2: Generalized Cost as a function of the value of time

A first remark is that the support of $\tau$ is invariant. This is a general property as shown in the previous section. Secondly the convexity of the travel time function increases with heterogeneity so that the travel time is higher in the centre of the peak period but lower on its flanks. At first sight it could lead to think that the aggregate cost is higher when heterogeneity is high. Figure 6.5 shows it is not the case: the aggregate cost decreases with heterogeneity. This gain is in fact due to a decrease in the costs incurred by the lowest VoT, while the cost incurs by the highest VoT is unaffected by the changes in heterogeneity.

In order to investigate this last phenomenon, the travel time costs and the schedule delay costs are plotted below. First no that in this case, the schedule delay costs are varying linearly with $\nu$. As it can be seen from Equation (6.11) and (6.3), this is solely due to the choice of a uniform distribution and it would be different otherwise. Second, for high heterogeneity, the travel costs are no longer monotously varying with the VoT and admit a maximum for a value of time $\bar{\nu} > \nu_m$. Finally, from the two figures it is clear that the decrease in the aggregate cost is solely due to a decrease in travel time costs and that the schedule delay costs remain unchanged.

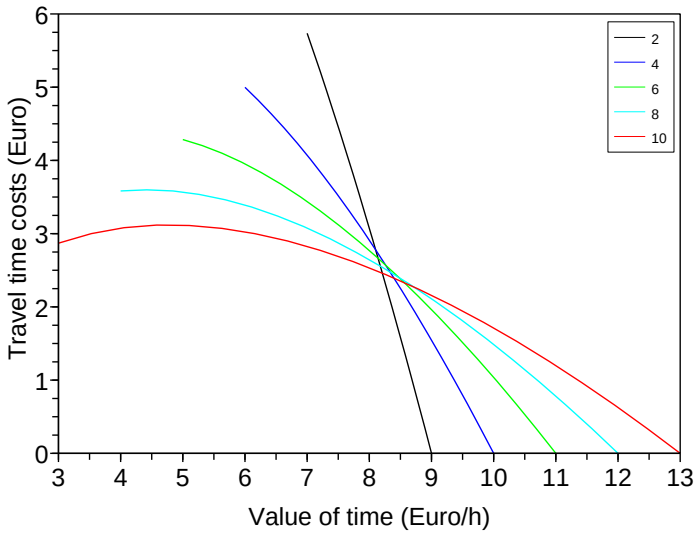



Figure 6.3: Travel time cost as a function of the value of time

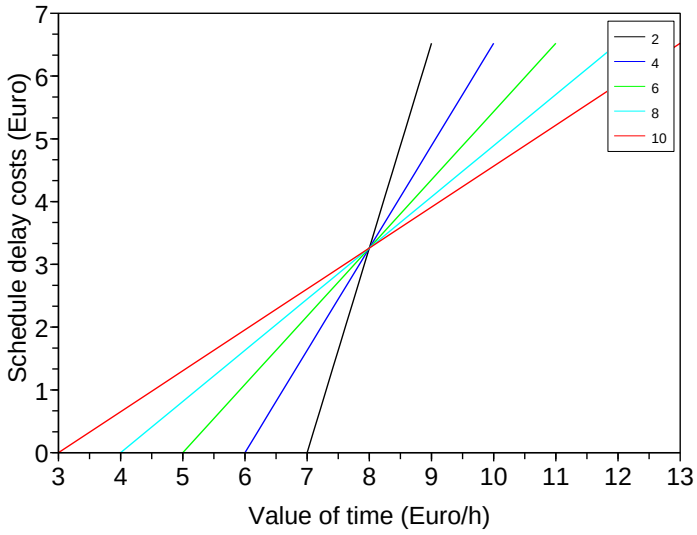



Figure 6.4: Schedule delay cost as a function of the value of time

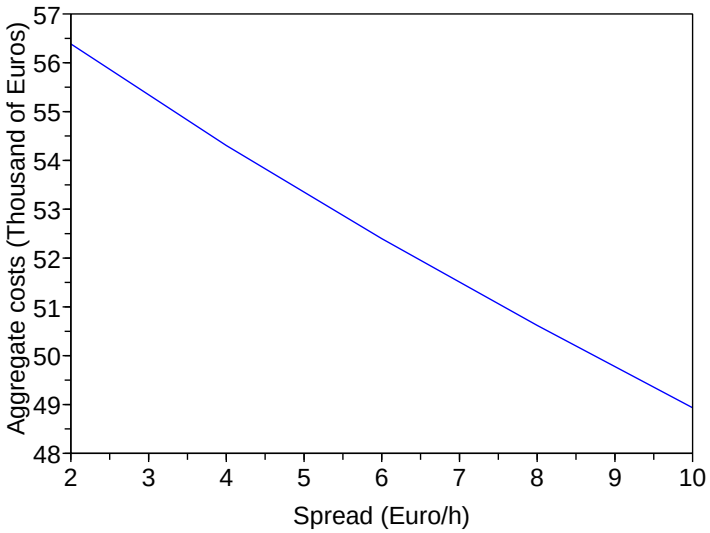



Figure 6.5: Aggregate Ccst as a function of the spread of the distribution

# 2 Extension with two routes

## 2.1 Model

We are going to define the extension to the two-route problem by analogy with the single route case. Consider a single OD pair served by two routes, 1 and 2, with respective capacity $(1-\rho).k$ and $\rho.k$. Both routes are priced at flat (*i.e.* time invariant) tolls $p_1$ and $p_2$ and have free flow travel times $t_1$ and $t_2$. As in this problem only the difference of costs on the two route matters, we assume without loss of generality that $p_1 = 0$ and $t_2 = 0$ and denote $p_2 = p$ and $t_1 = t > 0$. With loss of generality, we assume $p_2 > 0$. We will say that route 1 is the untolled route and that route 2 is the tolled route.

**Demand** In the two route case, an assignment of the demand has two components: the assignment of users between the two routes and the assignment between the arrival times within each route.

The users route decision is modelled by two values of time $\nu_1^\star$ and $\nu_2^\star$ such that $\nu_2^\star < \nu_1^\star$. The interval $[\nu_m, \nu_1^\star]$ corresponds to the VoT of users patronizing route 1, while $[\nu_2^\star, \nu_M]$ corresponds to the VoT of users patronizing

route 2. Remark that this description of demand is very specific; the implicit assumptions it carries will be discussed in the next subsection.

The arrival time choice is described by two pairs of functions, $(h^i_-, h^i_+)_{i\in\{1;2\}}$, as it was done in the model with one route. For both routes, the functions $(h^i_-, h^i_+)_{i\in\{1;2\}}$ are continuously differentiable and are respectively increasing and decreasing. Moreover it is assumed that $h^1_-(\nu_m) = h^1_+(\nu_m) = 0$ and $h^2_-(\nu^\star_2) = h^2_+(\nu^\star_2) = 0$. To be consistent with the users route decision, the functions $(h^1_-, h^1_+)$ should be defined on $[\nu_m, \nu^\star_1]$ and the functions $(h^2_-, h^2_+)$ on $[\nu^\star_2, \nu_M]$. To simplify the analytics, let us extend them continuously on $[\nu_m, \nu_M]$ by setting $h^1_-(\nu) = h^1_-(\nu_1)$ and $h^1_+(\nu) = h^1_+(\nu_1)$ on $[\nu^\star_1, \nu_M]$, and $h^2_-(\nu) = h^2_-(\nu_2) = 0$ and $h^2_-(\nu) = h^2_-(\nu_2) = 0$ on $[\nu_m, \nu^\star_1]$.

A state of the demand is thus represented by two triples $\Theta_1 = (h^1_-, h^1_+, \nu^\star_1)$ and $\Theta_2 = (h^2_-, h^2_+, \nu^\star_2)$. Consider given a state of the demand, it is then possible to define the patronage of each route conditional to $\Theta_1$ and $\Theta_2$:

$$N_1(\nu; \Theta_1) := (1-\rho)k.(h^1_+(\nu) - h^1_-(\nu)) \tag{6.13}$$

$$N_2(\nu; \Theta_2) := \rho k.(h^2_+(\nu) - h^2_-(\nu)) \tag{6.14}$$

The functions $N_1$ and $N_2$ have the following interpretation: $N_i(\nu; \theta_i)$ represents the volume of users with a VoT under $\nu$ that patronize route $i$. Note that $N_i$ can then be interpreted as the VoT distribution of users patronizing route $i$.

**Supply** The travel time on each route is assumed to follow the standard pointwise bottleneck model. The state of supply is represented by two travel times functions, $(\tau_i)_{i\in\{1;2\}}$. To be consistent with a state of the demand, a function $\tau_i$ needs to be $N_i(.; \Theta_i)$-feasible.

**Additional notations** To state the model in a concise manner, let us introduce some additional notations. Consider given triples $\Theta_i = (h^i_-, h^i_+, \nu^i_\star)$ and $\tau_i$ for $i \in \{1; 2\}$. As in the previous section, we introduce:

$$g_i(h; \tau_i, \nu) := \nu\tau_i(h) + \alpha.(h - h_p)^- + \beta.(h - h_p)^+ + \nu t_i + p_i$$

and

$$\tilde{g}_i(\nu; \Theta_i, \tau_i) := \nu.\tau_i(h^i_-(\nu)) + \alpha.(h^i_-(\nu) - h_p) + \nu t_i + p_i$$

$\tilde{g}_i(\nu; \Theta_i, \tau_i)$ expresses the cost incurred by a user with value of time $\nu$ undertaking route $i$ and choosing to arrive at $h^i_-(\nu)$.

**User equilibrium statement** We are now ready to generalize Definition 6.1 in the two-route case.

**Definition 6.7** (Dynamic User Equilibrium problem with two routes (DUE2R))**.** *Find a state of the demand* $(\Theta_i)_{i\in\{1;2\}} = (h^i_-, h^i_+, \nu^i_\star)_{i\in\{1;2\}}$ *together with two travel time functions* $(\tau_i)_{i\in\{1;2\}}$ *such that:*

1. *The triple* $(h^1_-, h^1_+, \tau_1)$ *is a solution to the DUE problem with one route of capacity* $(1-\rho)k$ *and a VoT distribution of* $N_1(.; \Theta_1)$*.*
2. *The triple* $(h^2_-, h^2_+, \tau_2)$ *is a solution to the DUE problem with one route of capacity* $\rho k$ *and a VoT distribution of* $N_2(.; \Theta_2)$*.*
3. *The following equations are satisfied:*

$$\min_i \tilde{g}_i(\nu; \Theta_i) = \begin{cases} \tilde{g}_1(\nu; \Theta_1, \tau_1) & on\ [\nu_m, \nu^\star_2] \\ \tilde{g}_1(\nu; \Theta_1, \tau_1) = \tilde{g}_2(\nu; \Theta_2, \tau_2) & on\ [\nu^\star_2, \nu^\star_1] \\ \tilde{g}_2(\nu; \Theta_2, \tau_2) & on\ [\nu^\star_1, \nu_M] \end{cases} \tag{6.15}$$

$$N_1(\nu; \Theta_1) + N_2(\nu; \Theta_2) = V(\nu) \tag{6.16}$$

The two first conditions express that the assignment of the demand represented by $(h^i_-, h^i_+)_{i\in\{1;2\}}$ is optimal within each route for a travel time $\tau_i$. In other words for a given value of time $\nu$, there is no best arrival time choice on the route $i$ than $h^i_-(\nu)$ and $h^i_+(\nu)$. Equation (6.15) expresses that $\nu^\star_1$ and $\nu^\star_2$ are consistent with an optimal assignment of the demand between the two routes. It is interesting to see that this statement of the problem is related to a formulation in a two stage decision problem: $h^i_-$ and $h^i_+$ encompass the optimal arrival time decision on each route, while $\nu^\star_1$ and $\nu^\star_2$ encompass the optimal route choice. Finally, Equation (6.16) is a volume conservation equation.

## 2.2 Comments about the equilibrium formulation

**Scope of the formulation** As for the single route case, the DUE problem, as presented here, states a specific equilibrium. The assumptions of the

previous model have been retained, so here again the travel time of each route admits a single peak centred on $h_p = 0$. Now users of each route needs to be at equilibrium w.r.t. their arrival time choice, so these assumptions can be justified by the same qualitative reasoning as for the one route model.

To obtain a simple route choice model, we introduced two critical VoT, namely $\nu_1^\star$ and $\nu_2^\star$. This assumption requires some explanations. If $\nu_1^\star = \nu_2^\star = \nu^\star$, it simply expresses that the low values of time patronize the untolled route while the high value of time prefer the tolled route, the limit being at certain critical value of time $\nu^\star$. Now, we also account for the existence of an intermediate category of users that patronize both routes, although possibly for different arrival times. In this latter setting there are two critical values of time $\nu_1^\star \neq \nu_2^\star$, dividing the transport demand among the two routes as illustrated in Figure 6.6. Accounting for this kind of equilibriums is critical, as we will see later that otherwise there might be cases where no equilibrium exists.

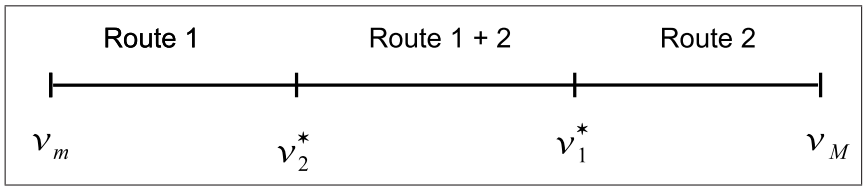



Figure 6.6: Illustration of the route choice model

**Influence of $t$ and $p$ on the equilibrium** In this extension to two routes, we introduced a non null free flow travel time on route 1 as well as a flat toll on route 2. Note that they do not appear in the two first conditions of Definition 6.7. This is natural: as they are both time-invariant, they have no impact in the arrival time choice of users on a given route. Now, they do influence users' route choice: the higher $p$ is the less attractive route 2 is; the higher $t$ is, the less attractive route 1 is. While deriving the solutions of the DUE2R problem, it will be shown that $\nu_1^\star$ and $\nu_2^\star$ depends of $p$ and $t$.

## 2.3 Derivation

**The two types of equilibriums** To derive the solutions of the DUE2R problem, it is easier to distinguish between two types of equilibriums. The first type of equilibrium is referred to as equilibriums of type $a$ and is such

that $\nu_1^\star \neq \nu_2^\star$. The second type of equilibrium is referred to as equilibriums of type $b$ and is such that $\nu_1^\star = \nu_2^\star$.

**The equilibrium cost problems for two routes** As in the problem with one route, the analytical resolution leads to consider second order differential equations in $\nu \to \tilde{g}(\nu)$, a quantity that can be interpreted as the cost incurred at equilibrium by a user with VoT $\nu$. Two equilibrium cost problems are introduced, one for equilibriums of type $a$, the other one for equilibriums of type $b$.

**Definition 6.8** (Equilibrium Cost Problem for equilibriums of type $a$ (ECP2Ra))**.** *Letting $\eta := \alpha\beta/(\alpha+\beta)\rho k$, solve the following differential equation on $[\nu_m, \nu_M]$:*

$$\frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu) = \begin{cases} -\dfrac{\eta v(\nu)}{\rho \nu} & \text{if } \nu > \nu_1^\star \\ -\dfrac{\eta v(\nu)}{\nu} & \text{if } \nu_2^\star < \nu < \nu_1^\star \\ -\dfrac{\eta v(\nu)}{(1-\rho)\nu} & \text{otherwise} \end{cases} \tag{6.17}$$

*where $\nu_1^\star$ and $\nu_2^\star$ are characterized by the following relationships:*

$$\tilde{g}(\nu_2^\star) = \eta V(\nu_2^\star) + \nu_2^\star t + \rho p$$

$$\eta V(\nu_1^\star) = (1-\rho).p$$

*and with the following boundary conditions:*

$$\frac{\partial \tilde{g}}{\partial \nu}(\nu_M) = 0$$

$$\tilde{g}(\nu_M) = \frac{\eta}{\rho}\Big(V(\nu_M) - V(\nu_2^\star)(1-\rho)\Big) + \rho p$$

Equation (6.17) describes a second order differential equation in $\tilde{g}$ that can be solved knowing $\nu_1^\star$ and $\nu_2^\star$ as well as boundary conditions. This equation is the same as in the one route case for $\nu$ in $[\nu_2^\star, \nu_1^\star]$. The second derivative of $\tilde{g}$ is lower for $\nu \notin [\nu_2^\star, \nu_1^\star]$, as a result of a lack of use of the total capacity of the system. This gives a hint regarding the non optimality of the equilibrium situation: a better use of the capacity would certainly results in lower total costs.

**Definition 6.9** (Equilibrium Cost Problem for equilibriums of type $b$ (ECP2Rb))**.** *Letting $\eta := \alpha\beta/(\alpha+\beta)\rho k$, solve the following differential equation on $[\nu_m, \nu_M]$:*

$$\frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu) = \begin{cases} -\dfrac{\eta v(\nu)}{\rho \nu} & \text{if } \nu > \nu^\star \\ \\ -\dfrac{\eta v(\nu)}{(1-\rho)\nu} & \text{otherwise} \end{cases} \tag{6.18}$$

*where $\nu^\star$ is defined by:*

$$\tilde{g}(\nu^\star) = \frac{\eta}{1-\rho} V(\nu^\star) + \nu^\star t$$

*and with the following boundary conditions:*

$$\frac{\partial \tilde{g}}{\partial \nu}(\nu_M) = 0$$

$$\tilde{g}(\nu_M) = p + \frac{\eta}{\rho}\big(V(\nu_M) - V(\nu^\star)\big)$$

There again the ECP2Rb is a second order differential equation which is similar to the one route case.

**Equivalency statement** In order to concisely state an equivalence between the ECP2Ra, the ECP2Rb and the DUE2R problem, it is first required to introduce the following definition.

**Definition 6.10** (Restatement of the DUE2R)**.** *Consider a continuous, convex and increasing function $\nu \to \tilde{g}(\nu)$ as well as two values of time $\nu_1^\star \geq \nu_2^\star$. Let the quantities $(h_+^i, h_-^i)_{i\in\{1;2\}}$ and $(\tau_i)_{i\in\{1;2\}}$ be defined from the equations below:*

$$\begin{aligned} h_-^1(\nu) &= -\frac{\tilde{g}(\nu_1^\star) - \nu_1^\star t}{\alpha} + \int_{\nu_1^\star}^{\nu} \frac{\nu}{\alpha} \frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu)\, d\nu \\ h_+^1(\nu) &= \frac{\tilde{g}(\nu_1^\star) - \nu_1^\star t}{\beta} - \int_{\nu_1^\star}^{\nu} \frac{\nu}{\beta} \frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu)\, d\nu \\ \tau_1\big(h_-^1(\nu)\big) &= \tau_1\big(h_+^1(\nu)\big) = \frac{\partial \tilde{g}}{\partial \nu}\big(\nu\big) - t \end{aligned}$$

*and*

$$
\begin{aligned}
h^2_-(\nu) &= -\frac{\tilde{g}(\nu_M)-p}{\alpha} + \int_{\nu_M}^{\nu} \frac{\nu}{\alpha}\frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu)\ d\nu \\
h^2_+(\nu) &= \frac{\tilde{g}(\nu_M)-p}{\beta} - \int_{\nu_M}^{\nu} \frac{\nu}{\beta}\frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu)\ d\nu \\
\tau_2\big(h^2_-(\nu)\big) &= \tau_2\big(h^2_+(\nu)\big) = \frac{\partial \tilde{g}}{\partial \nu}\big(\nu\big)
\end{aligned}
$$

*The triple* $(\tilde{g}, \nu_1^\star, \nu_2^\star)$ *is said* to solve the DUE2R problem *if and only if the state of the demand* $(\Theta_i)_{i\in\{1;2\}} = (h^i_+, h^i_-, \nu_i^\star)_{i\in\{1;2\}}$ *and the two travel times* $(\tau_i)_{i\in\{1;2\}}$ *does.*

Although its formulation is rather tedious, this definition is expressing a simple idea: it is restating the DUE2R with the triple $(\tilde{g}(\nu), \nu_1^\star, \nu_2^\star)$ as the base variable. Note that the equations stated in the definition have no specific justifications yet. They should be taken for granted for now and will make sense in the proof of the Proposition 6.11.

Let us now state how the problems ECP2Ra and ECP2Rb are equivalent to the DUE2R problem.

**Proposition 6.11** (Equivalency of the DUE2R and the two ECP2R)**.**

*(1) If the two quadruples* $\Theta_i = (h^i_-, h^i_+, \tau_i, \nu_i^\star)$*, for* $i \in \{1;2\}$*, solves the DUE2R problem, then* $\tilde{g} := \min_i \tilde{g}_i(.;\Theta_i)$ *solves either the ECP2Ra (and then* $\nu_1^\star \neq \nu_2^\star$*) or the ECP2Rb (and then* $\nu_1^\star = \nu_2^\star$*).*

*(2) Consider* $(\tilde{g}_a, \nu_1^\star, \nu_2^\star)$ *and* $(\tilde{g}_b, \nu^\star)$ *the respective solutions of the ECP2Ra and the ECP2Rb. Then:*

- *(i)* $\nu_1^\star = \nu_2^\star \quad \Rightarrow \nu^\star = \nu_1^\star = \nu_2^\star$ *and* $\tilde{g}_a = \tilde{g}_b$ *are solutions to the DUE2R problem.*
- *(ii)* $\nu_1^\star < \nu_2^\star \quad \Rightarrow (\tilde{g}_a, \nu_1^\star, \nu_2^\star)$ *is a solution of the DUE2R problem.*
- *(iii)* $\nu_1^\star > \nu_2^\star \quad \Rightarrow (\tilde{g}_b, \nu^\star, \nu^\star)$ *is a solution of the DUE2R problem.*

Proposition 6.11 shows that the DUE2R problem is either equivalent to the ECP2Ra or the ECP2Rb, depending on the exogenous variables of the problem.

## 2.4 General properties of the DUE

As in the one route case, the equivalency result of Proposition 6.11 directly yields some general properties of the user equilibrium. The two first ones are the same as for the one route DUE.

**Property 6.12** (Existence and uniqueness of the DUE)**.** *There exists a unique DUE as expressed in Definition 6.7.*

This is a direct consequence of Proposition 6.11 and of the existence and uniqueness of the solutions of the ECP2Ra and the ECP2Rb.As both equilibrium cost problems have a unique solution, also has the DUE2R problem. This highlights the importance of formulating the DUE2R so that equilibriums of both types, $a$ and $b$, might be considered. Not formulated as such, there would not be any guarantee of existence of DUE2R and the problem would thus be ill-posed.

**Property 6.13** (On the equilibrium cost properties)**.** *The function $\tilde{g}$ is convex and increasing.*

As in the one route case the total cost incurred by a user increases with his value of time.

**Property 6.14** (On the travel times of users sharing the same value of time)**.** *At equilibrium, two users with the same value of time incur the same travel times.*

This is a rather surprising property: although they might use a different route and arrive at a different time, users with the same value of time experience the same travel time. This latter case happens for equilibriums of type $a$ for users with VoT between $\nu_2^\star$ and $\nu_1^\star$. The toll is then only compensated by a reduction in the schedule delay costs: at equilibrium, users of the tolled route arrive closer to their preferred time than their best alternative on the untolled route, but they have to pay for this privilege.

Property 6.14 allows predicting the general travel time pattern that will be observed at equilibrium. This is depicted in Figure 6.7 for an equilibrium of type $a$. From this figure, it is easy to understand how one switches from an equilibrium of type $a$ to a one of type $b$. When $t$ and $p$ are high enough the red part parts of the travel times curves completely disappear, leading to a total segregation between users with high VoT and the ones with low VoT.

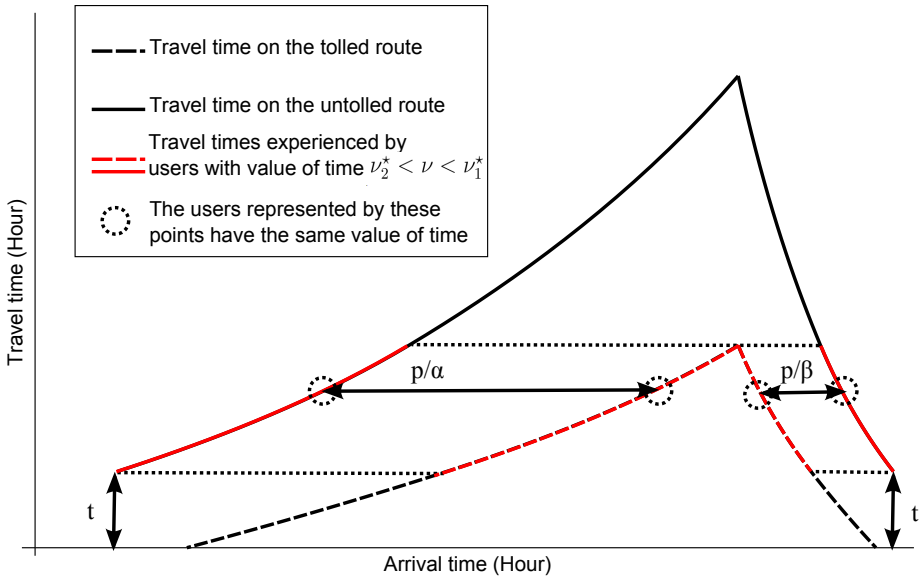



Figure 6.7: Travel time pattern in an equilibrium of type $a$

## 2.5 Proof of the equivalency result

The three following paragraphs give the proof of Proposition 6.11. The two first ones deals with the part (1) of the proposition, while the last one deals with the part (2). The proof is essentially an adaptation of the one of Proposition 6.3. The most technical parts of the proof have been transferred to the Appendix B for the sake of readability.

**Necessary conditions for equilibriums of type *a*** Assume given $(\Theta_i)_{i\in\{1,2\}} = (h_-^i, h_+^i, \nu_\star^i)_{i\in\{1,2\}}$ and $(\tau)_{i\in\{1,2\}}$, a solution to the DUE2R. Assume moreover that this solution is an equilibrium of type $a$.

Then denote $\tilde{g}(\nu) = \min_i \tilde{g}_i(\nu;\Theta_i)$. By definition of a DUE2R:

$$\tilde{g}(\nu) = \begin{cases} \tilde{g}_1(\nu;\Theta_1,\tau_1) & \text{for } \nu \text{ in } [\nu_m, \nu_1^\star] \\ \tilde{g}_1(\nu;\Theta_1,\tau_1) = \tilde{g}_2(\nu;\Theta_2,\tau_2) & \text{for } \nu \text{ in } [\nu_1^\star, \nu_2^\star] \\ \tilde{g}_2(\nu;\Theta_2,\tau_2) & \text{for } \nu \text{ in } [\nu_2^\star, \nu_M] \end{cases} \tag{6.19}$$

As in the previous section, let us first write the first order conditions of

optimality on $g_i$:

$$\frac{\partial g_i}{\partial h}\big(h^i_+(\nu);\Theta_i,\tau_i\big) = \frac{\partial g_i}{\partial h}\big(h^i_-(\nu);\Theta_i,\tau_i\big) = 0 \tag{6.20}$$

Consequently:

$$\begin{cases} \dfrac{\partial \tau_i}{\partial h}(h^i_-(\nu)) &= \dfrac{\alpha}{\nu} \\ \\ \dfrac{\partial \tau_i}{\partial h}(h^i_+(\nu)) &= -\dfrac{\beta}{\nu} \end{cases} \tag{6.21}$$

Equations (6.21) and (6.16) can be used to derive an equation on $\tilde{g}(\nu)$. Indeed:

$$\frac{\partial \tilde{g}}{\partial \nu}(\nu) = \begin{cases} \tau_1\big(h^1_-(\nu)\big) &= \tau_1\big(h^1_+(\nu)\big) & \text{for } \nu \text{ in } [\nu_m, \nu^\star_2] \\ \tau_1\big(h^1_-(\nu)\big) + t &= \tau_2\big(h^2_-(\nu)\big) & \text{for } \nu \text{ in } [\nu^\star_2, \nu^\star_1] \\ \tau_1\big(h^1_+(\nu)\big) + t &= \tau_2\big(h^2_+(\nu)\big) & \text{for } \nu \text{ in } [\nu^\star_2, \nu^\star_1] \\ \tau_2\big(h^2_-(\nu)\big) &= \tau_2\big(h^2_+(\nu)\big) & \text{for } \nu \text{ in } [\nu^\star_1, \nu_M] \end{cases} \tag{6.22}$$

The Equation (6.17) on the second derivative of $\tilde{g}$ can be derived from (6.22) and (6.16).

$$\frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu) = \begin{cases} -\dfrac{\alpha\beta}{(\alpha+\beta)\rho k}.\dfrac{v(\nu)}{\nu} & \text{if } \nu > \nu^\star_1 \\ \\ -\dfrac{\alpha\beta}{(\alpha+\beta)k}.\dfrac{v(\nu)}{\nu} & \text{if } \nu^\star_2 < \nu < \nu^\star_1 \\ \\ -\dfrac{\alpha\beta}{(\alpha+\beta)(1-\rho)k}.\dfrac{v(\nu)}{\nu} & \text{otherwise} \end{cases}$$

It remains to identify $\nu^\star_1$ and $\nu^\star_2$ as well as the boundary conditions. They are given by the two following lemmas.

**Lemma 6.15.** *Assuming a DUE of type a, $\nu^\star_1$ and $\nu^\star_2$ are characterized by the following relationships:*

$$\tilde{g}(\nu^\star_1) = \eta V(\nu^\star_1) + \nu^\star_1 t + \rho p$$

*and*

$$\eta V(\nu^\star_2) = (1-\rho).p$$

**Lemma 6.16.** *The boundary conditions are:*

$$\frac{\partial \tilde{g}}{\partial \nu}(\nu_M) = 0$$

$$\tilde{g}(\nu_M) = \frac{\eta}{\rho}\Big(V(\nu_M) - V(\nu_1^\star)(1-\rho)\Big) + \nu_1^\star t + \rho p$$

The proofs of Lemma 6.15 and 6.16 are given in Appendix B.

**Necessary conditions for equilibriums of type *b*** Assume given $(\Theta_i)_{i\in\{1,2\}} = (h_-^i, h_+^i, \nu_\star^i)_{i\in\{1,2\}}$ and $(\tau)_{i\in\{1,2\}}$, a solution to the DUE2R. Assume moreover that this solution is an equilibrium of type $b$.

Then denote $\tilde{g}(\nu) = \min_i \tilde{g}_i(\nu;\Theta_i)$ and $\nu^\star = \nu_1^\star = \nu_2^\star$. Expressing the first order conditions eventually gives:

$$\frac{\partial \tilde{g}}{\partial \nu}(\nu) = \left\{ \begin{array}{rcll} \tau_1(h_-^1(\nu)) & = & \tau_1(h_+^1(\nu)) & \text{for } \nu \text{ in } [\nu_m, \nu^\star] \\ \tau_2(h_-^2(\nu)) & = & \tau_2(h_+^2(\nu)) & \text{for } \nu \text{ in } [\nu^\star, \nu_M] \end{array} \right. \tag{6.23}$$

The two properties expressed in the previous Subsection are thus also true for equilibriums of type $b$. The equation on the second derivative of $\tilde{g}$ is given by:

$$\frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu) = \left\{ \begin{array}{ll} -\dfrac{\alpha\beta}{(\alpha+\beta)\rho k}.\dfrac{v(\nu)}{\nu} & \text{if } \nu > \nu^\star \\ \\ -\dfrac{\alpha\beta}{(\alpha+\beta)(1-\rho) k}.\dfrac{v(\nu)}{\nu} & \text{otherwise} \end{array} \right. \tag{6.24}$$

Finally $\nu^\star$ and the boundary conditions are given by the two following Lemmas.

**Lemma 6.17.** *Assuming a DUE of type b, $\nu^\star$ is characterized by the following relationship:*

$$\tilde{g}(\nu^\star) = \frac{\eta}{1-\rho} V(\nu^\star) + \nu^\star t$$

**Lemma 6.18.** *The boundary conditions are:*

$$\frac{\partial \tilde{g}}{\partial \nu}(\nu_M) = 0$$

$$\tilde{g}(\nu_M) = p + \frac{\eta}{\rho}\big(V(\nu_M) - V(\nu^\star)\big)$$

The proofs of Lemma 6.17 and 6.18 are similar to the one of Lemma 6.15 and Lemma 6.16.

**Sufficient conditions for the equilibrium cost problem with two routes** The previous paragraphs have exhibited necessary conditions for the equilibrium under for two different types of equilibriums, $a$ and $b$. In both case they describe a unique candidate solution as a solution of a second order differential equation. It remains to be shown that for a given set of parameters, there is always a unique valid solution among the two candidates. The proof of the "sufficiency" part of Proposition 6.11 is given in Appendix B.

## 2.6 Analytical example with a uniform distribution

Let us now assume a distribution $V$ in order to get some insights regarding the user repartition between the untolled route and the tolled route. In the following, we set $V(v) = \theta.(\nu - \nu_m)$, hence choosing a uniform distribution of the values of time with density $\theta$. Assuming an equilibrium of type $a$ and with $\nu_1^\star$ and $\nu_2^\star$ known, the integration is straightforward:

$$\tilde{g}(\nu) = \begin{cases} \dfrac{\theta\eta\nu}{\rho}\ln(\nu/\nu_M) + \dfrac{\theta\eta}{\rho}(\nu - \nu_M) + \tilde{g}(\nu_M) & \text{if } \nu > \nu_2^\star \\ \theta\eta\nu\ln(\nu/\nu_2^\star) + (\theta\eta + t)(\nu - \nu_2^\star) + \tilde{g}(\nu_2^\star) & \text{if } \nu_1^\star < \nu < \nu_2^\star \\ \dfrac{\theta\eta\nu}{1-\rho}\ln(\nu/\nu_1^\star) + \dfrac{\theta\eta}{1-\rho}(\nu - \nu_1^\star) + \tilde{g}(\nu_1^\star) & \text{otherwise} \end{cases}$$

Similarly, assuming an equilibrium of type $b$:

$$\tilde{g}(\nu) = \begin{cases} \dfrac{\theta\eta\nu}{\rho}\ln(\nu/\nu_M) + \dfrac{\theta\eta}{\rho}(\nu - \nu_M) + \tilde{g}(\nu_M) & \text{if } \nu > \nu_2^\star \\ \dfrac{\theta\eta\nu}{1-\rho}\ln(\nu/\nu^\star) + \dfrac{\theta\eta}{1-\rho}(\nu - \nu^\star) + \tilde{g}(\nu^\star) & \text{otherwise} \end{cases}$$

It remains to explicit the expressions of $\nu_1^\star$ and $\nu_2^\star$ with respect to the parameters of the problem. There are given by:

$$\nu_1^\star = \nu_m + (1-\rho)p/\theta\eta$$

$$\nu_2^\star = \nu_M e^{-\rho t/\theta\eta}$$

While $\nu^\star$ can only be expressed by the implicit equation:

$$\nu^\star \frac{\theta\eta}{\rho} \ln\left(\frac{\nu^\star}{\nu_M}\right) + \nu^\star t + \frac{\eta\theta}{1-\rho}(\nu^\star - \nu_m) = p$$

In the two following alineas, some numerical illustrations are given. The common numerical setting is presented in Table 6.1.

| Variable | Value |
|---|---|
| Capacity | 1800 users/hour |
| % of the capacity assigned to route 2 | 30% |
| Average value of time | 8€ |
| Toll | 5€ |
| Free flow travel time | 0.7 h |
| Nb of users | 3000 |
| $[\nu_m, \nu_M]$ | $[3, 13]$ |
| $\alpha$ | 4 €/h |
| $\beta$ | 16 €/h |

Table 6.1: Numerical values for the illustration

**Influence of the parameters on the equilibrium type** The expressions of $\nu_1^\star$, $\nu_2^\star$ and $\nu^\star$ allow us to study how the parameters $p$ and $t$ lead to one of the two possible equilibriums. Figure 6.8 depicts the distribution of the patronage between the two possible routes for $t = 0.7$ hour and a toll varying between 0 and 12 €. Figure 6.9 depicts the same situation except the toll is fixed at 5 €and $t$ is varying between 0 and 1.2 hours.

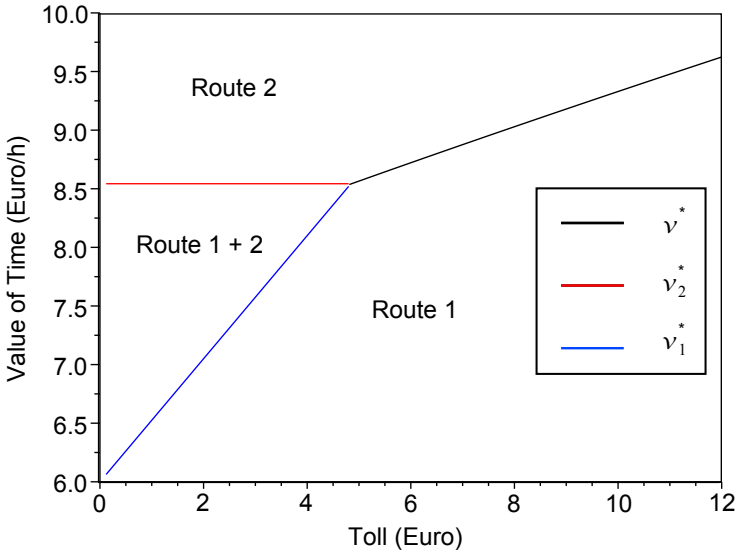



Figure 6.8: Critical values of times for different values of the toll

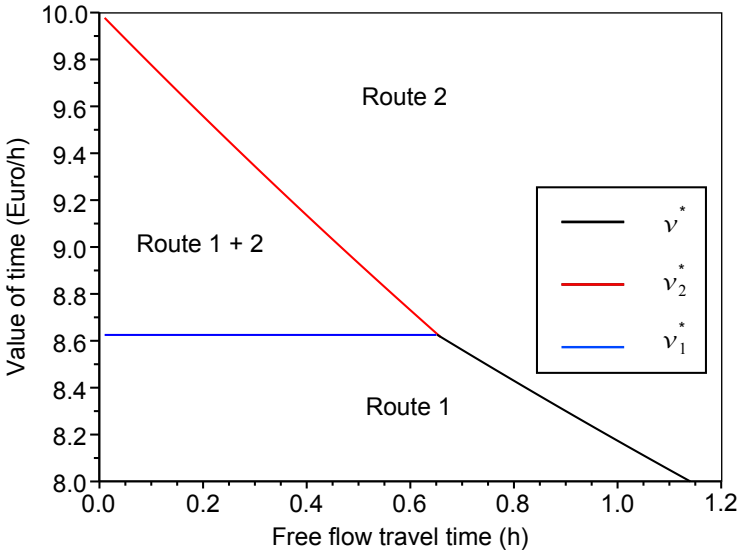



Figure 6.9: Critical values of times variation with the difference of free flow travel times

Finally, Figure 6.10 shows the partition of the space of parameters $(t, p)$ according to the two types of equilibriums of the system. The interpretation is as follow: when the differentiation between the two routes is low in terms of price as well as in term of free flow travel times an important part of the users patronize both routes. The observed equilibrium is of type $a$. When the two routes are more differentiated, the system switches to an equilibrium of type $b$ and there is a total segregation between the users with high VoT, patronizing the tolled route, while users with low VoT use the untolled route.

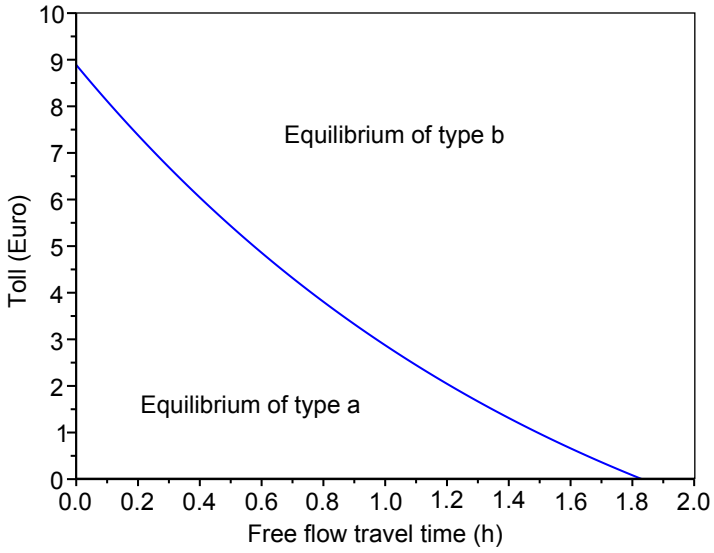



Figure 6.10: Equilibrium type according the parameters

**Influence of the VoT heterogeneity on the equilibrium** Let us study the impact of the VoT heterogeneity on the travel time and on the costs incur by the users. Two cases are considered. In the first one, a low toll of $3e$ is set which leads to an equilibrium of type $a$. In the second one, a high toll of $6e$ and an equilibrium type $b$ is observed.

Let us first deal with the *low toll case*. Figure 6.11a shows the generalized costs $\tilde{g}$ as a function of the value of time. , Figure 6.11b and Figure 6.11c show respectively the travel time as a function of the exit time for the two routes. Several values of $\theta = \nu_M - \nu_m$ have been tested. Globally the impact of an increase in heterogeneity is similar as in the one-route case: the peak tends

to concentrate on both routes. Yet an important difference can be observed: the support of the travel time is now varying. On the untolled route the more heterogeneous the user population is the shorter is the congested period. On the congested route, the contrary happens. This is related to the fact that the generalized cost of the user with the highest VoT is no longer a constant, as it was in the one-route case, but it is increasing with user heterogeneity.

Let us now expose the *high toll case*. As the travel time patterns are relatively similar to the low toll case, only the user costs are presented. Figure 6.12a, Figure 6.12c and Figure 6.12d depict the generalized costs, the travel time costs, and the schedule delay costs w.r.t. the VoT. Figure 6.11c is especially interesting as it shows a strong discontinuity in the travel time costs. Also note that in this case the generalized cost is decreasing for all VoT *i.e.* when heterogeneity is increasing, everybody is better off.

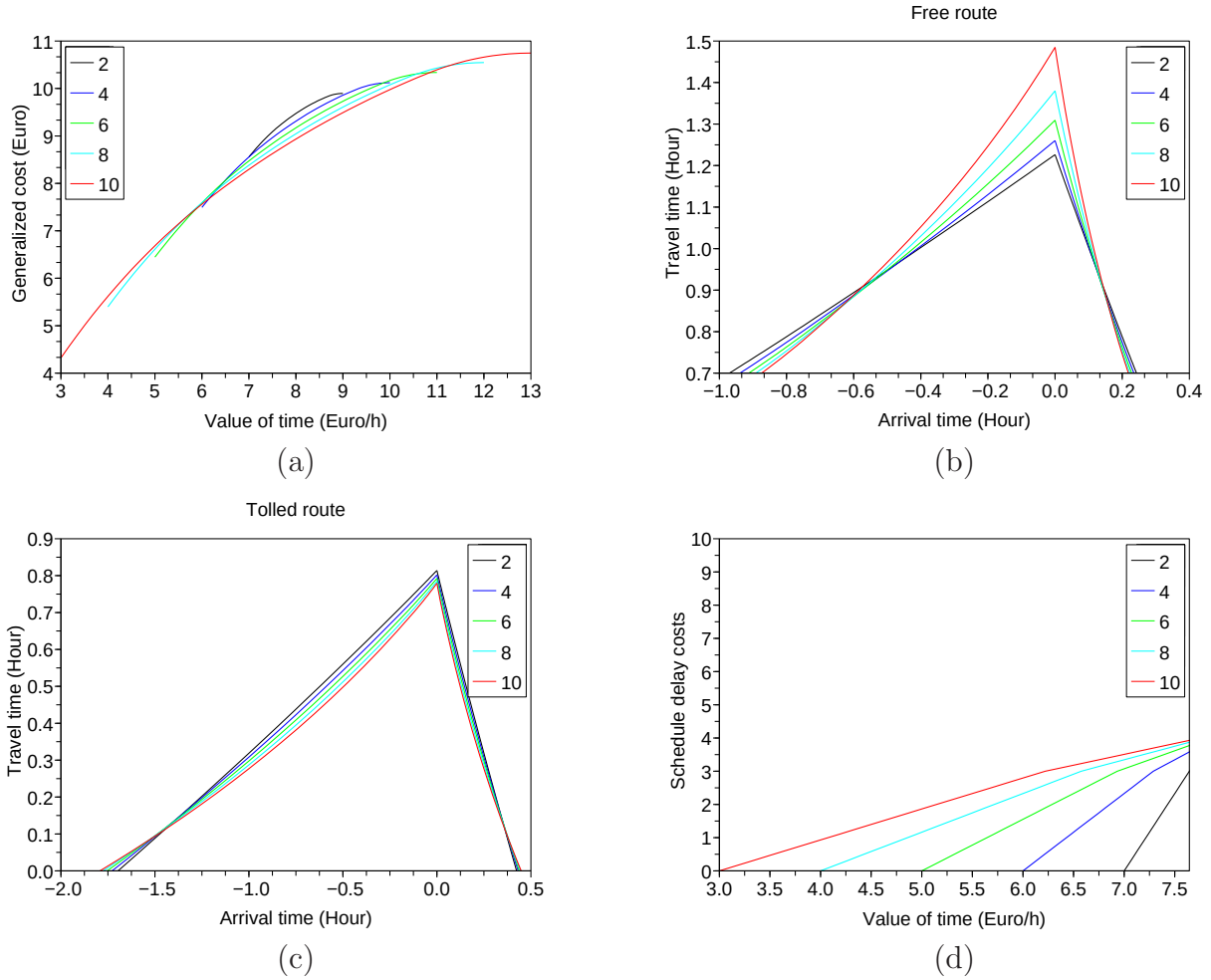



Figure 6.11: For different values of the spread in the low toll case: (a) generalized costs, (b) travel times on the untolled route, (c) travel times on the tolled route, and (d) schedule delay costs

Figure 6.12: For different values of the spread in the high toll case: (a) generalized costs for different values of the spread in the high toll case, (b) travel costs, (c) travel times, and (d) schedule delay costs

# 3 Pricing under two ownership regimes

We can now return to our initial objective. How is the relative efficiency of public and private pricing affected by user heterogeneity? In other words are there cases where the tolled route can be operate by a private company without public control and still guarantee a reasonable economic efficiency from a collective perspective? With the analytical model developed earlier, the answer to that question can now be precisely established.

Before going into the analysis, let us first define the tolling policies that are examined. In the two main policies that are tested, a flat toll (*i.e.* a time-invariant toll) is set on the tolled route. In the first policy, the toll is set to minimise of users costs minus the toll revenue, *i.e.* the social cost. Note that this first policy is equivalent to maximize welfare, as here no demand elasticity is introduced. In the second policy, the toll is set to maximise the toll revenue as if a private operator was setting it. In addition to those two policies, the no toll policy will be considered in order to give a benchmark. Table 6.2 summarizes the studied policies.

| Abbreviation | Description |
|---|---|
| NT | No toll |
| PBTI | Flat toll set by a public operator |
| PRTI | Flat toll set by a private operator |

Table 6.2: Abbreviation of the analysed policies

For analytical as well as numerical simplicity only uniform distributions of the VoT will be considered in the following. The no toll DUE is then computed directly from the formulas of the previous section. Both policies PBTI and PRTI requires the resolution of an optimization program to find the two corresponding flat tolls. They are computed numerically with a simple grid search. This search is helped by the fact that social costs and profits seem globally concave for all values of the time-invariant toll that have been tried.

The relative efficiency of the PRTI policy compared to the PBTI policy will be evaluated using the index:

$$\omega := \frac{c^{NT} - c^{PRTI}}{c^{NT} - c^{PBTI}}$$

where $c^X$ is the social cost of policy $X$. A similar index has previously been used for similar studies (among others in Arnott, de Palma and Lindsey (1991) or Verhoef, Nijkamp and Rietveld (1996)).

The section is structured as follows. First, numerical results summarizing the impact of different policies are analysed. Then, a sensitivity analysis is conducted on relevant parameters, namely $\rho$, $t$ and $k$. Finally, the impact of user heterogeneity w.r.t. the VoT is studied specifically. The numerical setting is the one presented in Table 6.1.

## 3.1 Numerical results for the different policies

Table 6.3 presents the results of the different policies. As expected the toll in the PRTI policy is significantly higher than in the PBTI case. This results in slightly higher travel times on the untolled route but much smaller travel times on the untolled route. Note that the efficiency of the PRTI policy is negative which implies it is less efficient than a no toll policy: in this case the private operator has no incentive to internalize part of the congestion costs of his users. This not always true and in some cases the PRTI policy might be efficient.

The difference in social costs between the PBTI and NT policies can be interpreted as such. When no toll is set, the route 2 (the "tolled" route) is highly congested and the schedule delay costs as well as the queuing costs are much higher than on route 1 (the "untolled" route). This is due to the high free flow travel times that are incurred on route 1. When a toll of 3.4 € is set on route 2, part of the users leave it in favour of route 1. As the external congestion costs are higher on route 2, this results in a global decrease of the queuing costs. The cost related to the free flow travel time naturally increases but it is compensated by a decrease in schedule delay costs. The switch from the NT policy to the PBTI policy thus leads to a decrease in social costs.

When it comes to the PRTI and PBTI policies, the higher toll set in the PRTI case still induce a slight decrease in the queuing costs and in the schedule delay costs. However, the increase in cost related to the free flow travel time is much more important. Thus the global cost increases and is even higher than in the no toll case.

| **Policy** | PRTI | PBTI | NT |
|---|---|---|---|
| Efficiency | -0.17 | 1 | 0 |
| User costs minus toll revenue (k€) | 26.6 | 25.1 | 26.4 |
| Travel time costs (k€) | 18.5 | 16.6 | 15.9 |
| Queuing costs (k€) | 8.0 | 8.5 | 10.5 |
| Schedule delay costs (k€) | 7.2 | 7.8 | 9.2 |
| Toll (€) | 7.15 | 3.45 | 0 |
| Average travel time on the untolled route (h) | 1.05 | 0.97 | 0.92 |
| Average travel time on the tolled route (h) | 0.25 | 0.48 | 0.71 |
| Patronage of the tolled route | 818 | 1263 | 1665 |

Table 6.3: Numerical results for the different policies

## 3.2 Sensivity analysis

**Sensivity to the difference in free flow travel times** Figure 6.13 shows the influence of the parameter $t$ on the index $\omega$. Two comments can be made.

First, the efficiency of the PRTI policy relatively to the efficiency of the PBTI policy is concavely increasing with $\tau$. Indeed when $\tau = 0$, the optimal public toll is 0, so the index $\omega$ is clearly $-\infty$. On the contrary the higher $\tau$ gets, the higher the optimal public toll should be, while the private toll is less affected by this change of parameters.

Second, note that the relative efficiency of the PRTI is significatively positive for high values of $\tau$, with an index $\omega$ close to 0.5. This is especially interesting from a policy perspective: when the difference in free flow travel time is important, a private operator acts as if it was partially internalizing the congestion costs of its users and thus prices his route in consequence. On the contrary for low values of $\tau$, the PRTI is especially $\omega$ close to 0.5 inefficient, with an index $\omega$ close to $-2.5$.

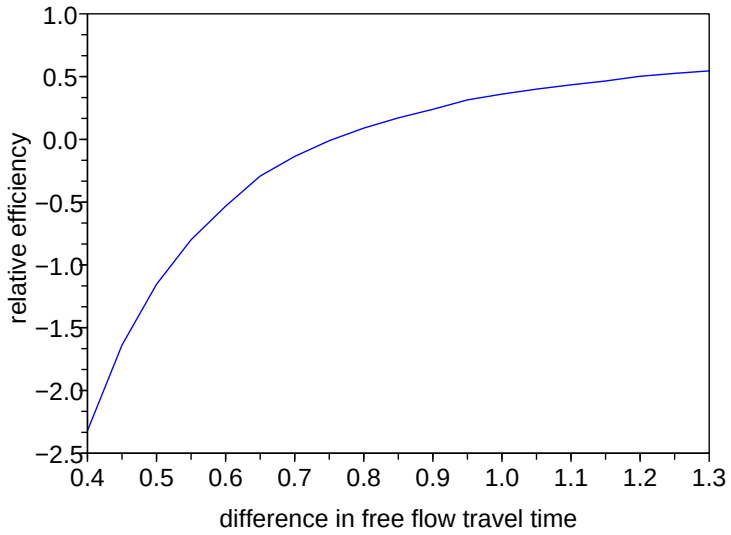



Figure 6.13: Sensibility of the relative efficiency of the PRTI policy to $\tau$

**Sensivity to the global capacity** The influence of $k$ on the relative efficiency of the PRTI policy has a simple pattern as shown by Figure 6.14. As in the previous case, the index $\omega$ is concavely increasing with $k$. Indeed, as $\rho = 0.3$ an increase in the global capacity strongly lower congestion on route 1 while it has fewer impacts on route 2. Consequently a private firm operating route 2 has to lower its toll in order to keep its patronage.

**Sensivity to the relative capacity between the two routes** Figure 6.15 shows that the index $\omega$ is extremely sensitive to $\rho$. The PRTI policy quickly decreases in efficiency. This is natural as for $\rho = 1$, there is no more alternative to the tolled route and thus the toll can be set as high as wanted by a private operator. This last comment is rather obvious, a more surprising fact is that even for moderately high values of $\rho$, the PRTI policy is dramatically inefficient: for $\rho = 0.5$ the index $\omega$ is already under $-3$.

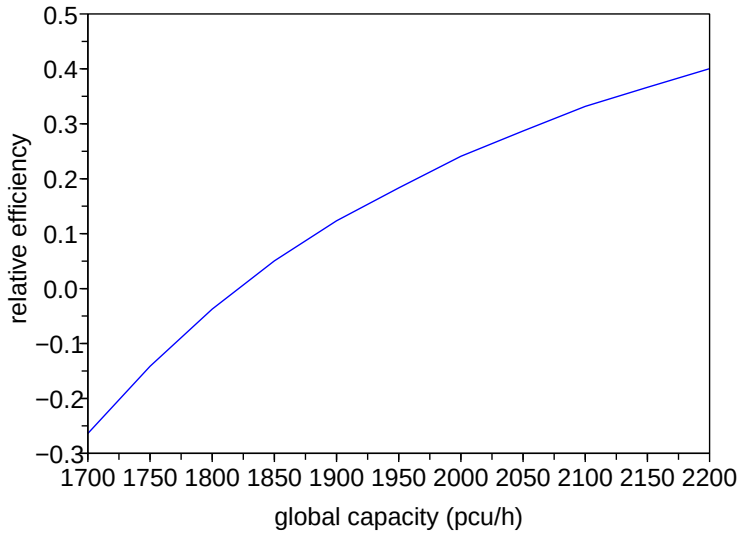



Figure 6.14: Sensibility of the relative efficiency of the PRTI policy to $k$

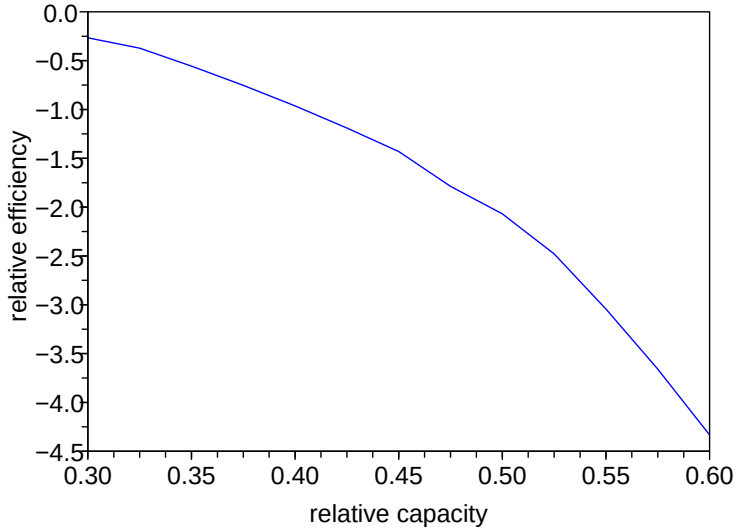



Figure 6.15: Sensibility of the relative efficiency of the PRTI policy to $\rho$

## 3.3 Impact of user heterogeneity on the relative efficiency of public and private pricing

Let us now investigate the impact of the user heterogeneity w.r.t. the VoT on the relative efficiencies of the policy. The intuitive result is that revenue maximizing policies should be more efficient with high heterogeneity. Indeed high heterogeneity favours such policies because product differentiation offers a greater advantage: those with high values of time reap more benefits from the high-priced option, while those with low values of time still enjoy the unpriced option. Such results are well known in static frameworks (see Small and Yan, 2001, for instance).

This intuition was proven wrong by van den Berg and Verhoef (2010) for time-varying toll policies. A possible interpretation is that as the level of heterogeneity in VoT grows the users naturally assign themselves more efficiently. This reduces the need for a toll-driven coordination.

The analytical model developed in this chapter allowed us to carry on this study with great precision for flat toll policies. Our main finding is that in fact the relative efficiencies of revenue maximizing policies are impacted positively by the heterogeneity in value of time. Numerical experiments lead us to think that this result is robust to changes in the numerical settings. Thus, the previous results established for static congestion seem to be still valid.

Figure 6.16 depicts this phenomenon for two values of $\tau$: the shape of each of the resulting curves is very similar. Note that when the difference in free flow travel time is high, the relative efficiency of the PRTI policy is more sensible to the level of heterogeneity in VoT. In this later case and for high values of the spread in VoT, the PRTI is nearly as efficient as the PBTI policy.

Now a closer look to the values reveals that on the whole the efficiency of the PRTI is rather low except when heterogeneity (the spread in value of time) and differentiation between the two routes (the difference in free flow travel time) is extreme. This was confirmed by numerous numerical tests.

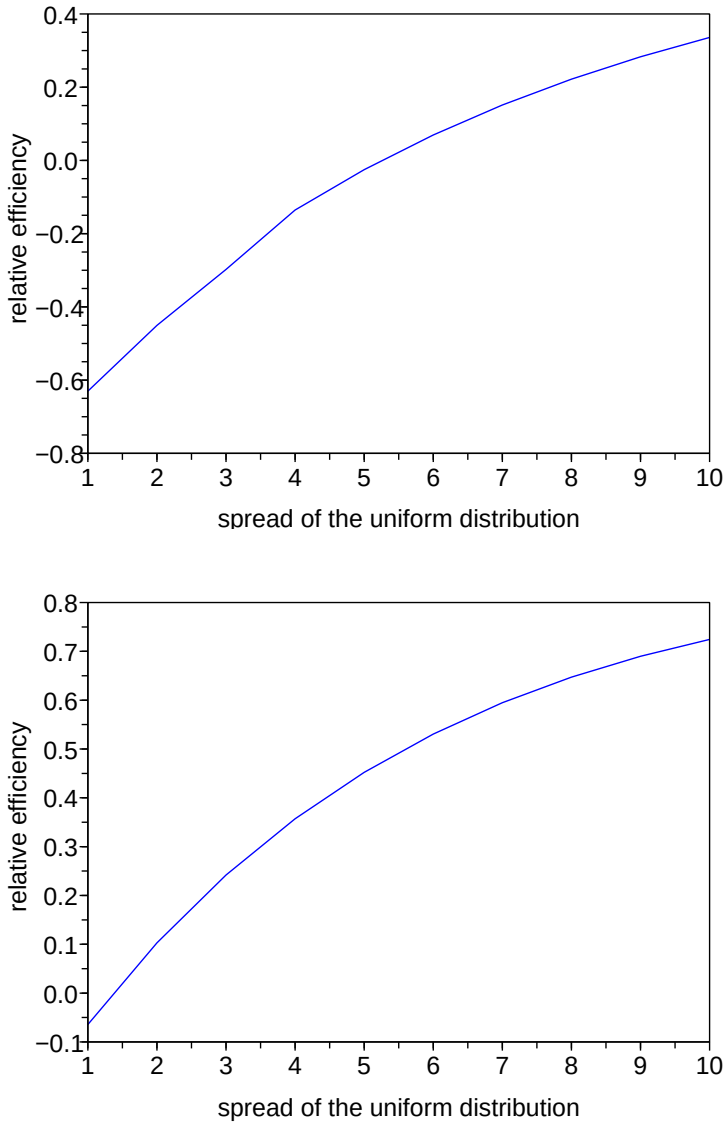



Figure 6.16: Sensibility of the relative efficiency of the PRTI policy to the spread of the VoT distribution for $\tau = 0.7$ (up) and $\tau = 1$ (down)

## Conclusion

This chapter demonstrates the importance of heterogeneity in value of time for evaluating congestion policies that offer pricing as an option. Our contributions are twofold.

First from a methodological point of view, a model that properly accounts for heterogeneity in VoT has been proposed and derived. General properties have been demonstrated. For instance at equilibrium, there might be users with the same VoT travelling on different routes although not at the same time ; contrary to the static case there is not necessarily a critical value of time dividing users between the two routes. Another surprising finding is that users with the same VoT always incur the same travel time and thus that the difference in costs induced by the toll is solely compensated by the schedule delay costs.

From a policy perspective, two strategies for flat toll pricing on the tolled route have been assessed. The first one is revenue maximizing while the second one is social cost minimizing. It has been shown that heterogeneity in VoT impacts positively the relative efficiency of the revenue maximizing policy. Consequently, when the heterogeneity is high, the social efficiency of a revenue maximizing policy can be very close to the one of a welfare maximizing policy. This is a known result for static congestion, and this chapter thus show it is still valid with dynamic congestion.

# Part IV

# Numerical resolution

Chapter 7

# A user equilibrium model with departure time choice

This chapter presents a model of Dynamic User Equilibrium (DUE) with departure time choice. The model is formulated in a supply-demand framework where the supply is a network of bottlenecks and the demand a set of microeconomic agents. Those agents are characterized by economic preferences (a value of time and a schedule delay cost function), a scheduling preference (a preferred arrival time) and physical characteristics (a vehicle class). The network is subject to congestion and tolled. Hence to a given demand the supply model associates time-varying travel time functions and toll functions with each route. Similarly demand reacts to supply by adjusting the time-varying flows at the entrance of each route of the network according to the level of congestion.

The main features of the model are:

1. time is represented continuously;
2. scheduling preferences are represented by continuous distributions of the preferred arrival times;
3. traffic flowing is multi-class;
4. time-varying tolls can be imposed on each arc.

In Chapter 10 we argue this set of assumptions is particularly well suited for interurban applications. The main technical technical difference between this model and the ones currently developed in the literature is that the trip scheduling model is deterministic. For examples of DUE models with

stochastic trip scheduling see for instance (De Palma and Marchal, 2002) or (Bellei et al., 2005).

## Objectives of the Chapter

In this chapter, our purpose is to present a DUE model compatible with the dynamic congestion games introduced in Chapter 3. In particular, the first formulation proposed allows users with the same characteristics to choose different departure times. This imposes a representation of the assignment of the transport demand, in the form of a joint distribution between the space of users and the space of travel decisions.

Now, from a computational perspective, this representation is inconvenient. Most of the existing models in the literature, rather assume that users' choices are *symmetric* with respect to their departure time *i.e.* where users with the same characteristics choose the same departure time. It is shown here that this representation is incorrect for a tolled network, as no equilibrium might exists with such a property. However a correct formulation is introduced, where users' choices are symmetric with respect to *their arrival time*.

# 1 A new formulation for the dynamic user equilibrium problem with departure time choice

In this model a set of users belonging to different categories wish to travel on a congestion-prone network. The travel time they experience on the network varies with their category, as they can drive vehicles of different classes (*e.g.* cars or trucks) and arcs on the network might be subject to time-varying tolls. Thus the level of a service of a route $r$ for a user of category $c$ leaving at $h$ is represented by a pair $(\tau_{rc}(h), p_{rc}(h))$ where $\tau_{rc}(h)$ is the travel time of the route when departing at $h$ and $p_{rc}(h)$ is the sum of the tolls along the route.

Let us first specify some technical details and notations.

**Simulation period** Denote $\mathcal{H} = [h_m, h_M]$ the simulation period, assumed to be "large enough" for all the trips to start and end in $\mathcal{H}$.

**Network** The sets $(A, N)$ models the network topology represented by a directed graph with nodes $n \in N$ and arcs $a \in A$. The set of $OD$ pairs is a subset of $N \times N$ and each $od \in O \times D$ is served by a set of routes $R_{od}$.

**Route flows** The flows are integrable functions and hence taken in $L_1(\mathcal{H}, \mathbb{R})$; they are denoted by lowcase $x$ and the associated uppercase symbol $X$ is the cumulated flow corresponding to $x$. The quantity $X$ can be seen either as a measure ($X(I)$ then counts the number of users passing during an interval $I$) – or as an increasing function – ($X(h)$ is then the number of users that have passed before $h$). The flows on the network are described by route cumulated flows of users from a given category $c$. There are denoted $X_{rc}$.

**Travel time and toll functions** Travel time and toll functions are taken in $\mathcal{C}(\mathcal{H}, \mathbb{R}_+^*)$, the set of continuous functions from $\mathcal{H}$ into $\mathbb{R}_+^*$.

**Vectors** Vector of travel time functions, toll functions and cumulated volumes are the main objects of this chapter. Let us adopt the following convention: for a quantity $X_{ij}$ subscripted with two variables $i$ and $j$ denote: $\boldsymbol{X}_{Ij} := (X_{ij})_{i \in I}$, $\boldsymbol{X}_{iJ} := (X_{ij})_{j \in J}$ and $\boldsymbol{X}_{IJ} := (X_{ij})_{i \in I, j \in I}$. In particular:

- the route flow vector is denoted $\boldsymbol{X}_{RC} := (X_{rc})_{r \in R, c \in C}$;
- the route flow vector of a user category $c$ is denoted $\boldsymbol{X}_{Rc} := (X_{rc})_{r \in R}$;
- the travel time functions and the toll functions vectors are denoted respectively by $\boldsymbol{\tau}_{RC} := (\tau_{rc})_{r \in R, c \in C}$ and $\boldsymbol{p}_{RC} := (p_{rc})_{r \in R, c \in C}$.

## 1.1 Transport Demand

**User model** Each user is characterized by:

- an OD pair $od$;
- a pair $e = (D, \nu)$ that represents his economic preferences. $D$ is a schedule delay cost function that associates to a delay $l$ ($l$ stands for lateness) his schedule cost. The quantity $\nu$ is a value of time;
- a vehicle class $u$;

- and $h_p$ his preferred arrival time.

Now consider a user $c = \{od, (D, \nu), u\}$ with arrival preference $h_p$. Assume he is leaving at an instant $h$ and incurring a travel time $\tau_{rc}(h)$ together with a monetary cost of $p_{rc}(h)$. In this context, one can write his *generalized travel cost* as:

$$g\big(h, \tau_{rc}(h), p_{rc}(h)|c, h_p\big) = \underbrace{\nu.\tau_{rc}(h) + p_{rc}(h)}_{\text{traversal costs}} + \underbrace{D\big(h_p - (h + \tau_{rc}(h))\big)}_{\text{schedule delay costs}} \quad (7.1)$$

Equation (7.1) states that the generalized travel cost is divided in two parts: (1) a traversal cost composed of the travel time costs and the tolls and (2) a schedule delay cost.

Users are assumed to be microeconomic agents seeking to minimize their generalized travel cost. Thus, given a set of route travel time functions $\boldsymbol{\tau}_{RC} = (\tau_{rc})_{r \in R_{od}}$ and tolls functions $\boldsymbol{p}_{RC}(p_{rc})_{r \in R_{od}}$, a user characterized by $(c, h_p)$ is solving the following program:

$$\min_{h,r} g\big(h, \tau_{rc}(h), p_{rc}(h)|c, h_p\big) \quad (7.2)$$

**User population** The economic preferences and the vehicle classes are taken respectively from the two finite sets $E = \{e_1, \ldots, e_{n_e}\}$ and $U = \{u_1, \ldots, u_{n_u}\}$. Arrival preferences are taken from a continuous interval $\mathcal{H}_p$ strictly included in $\mathcal{H}$. For technical reasons let us group all the discrete characteristics in one single set $C = N^2 \times E \times U$. The elements $c$ of $C$ are said to be *user categories*. Note that a pair $(c, h_p)$ fully characterizes a user.

The transport demand can be represented by collections (one for each user category) of cumulative distributions $X_c^p$ over $\mathcal{H}_p$. The quantity $X_c^p(h)$ represents the volume of users of category $c$ leaving from $o$ that would prefer to arrive at $d$ before $h$. The distribution $X_c^p$ is called *the distribution of preferred arrival times of category c*.

In our approach users are thus represented by a sequence of continuums of agents, one for each category. Each distribution $X_c^p$ represents one of these continuums.

**User travel choices** Assume $\boldsymbol{p}_{RC}$ and $\boldsymbol{\tau}_{RC}$ given. Each user belonging to category $c$ has to choose a route among the possible routes of the network $R$ and a departure time in $\mathcal{H}$. Hence their possible travel decisions lie in

$\mathcal{S} = \mathcal{H} \times R$. Choices of the users belonging to a category $c$ are represented by a measure $D_c$ on $\mathcal{H}_p \times \mathcal{S}$ such that the marginal[1] of $D_c$ on $\mathcal{H}_p$ is $X_c^p$. The distribution $D_c$ can be seen as an assignment of the demand represented by $X_c^p$ on the network. It will be referred to as a *departure distribution.*

One might wonder what is the relationship the distribution $D_c$ with more physical quantities such as the route cumulated flows. Denote $X_c$ the marginal of $D_c$ over $\mathcal{S} = \mathcal{H} \times R$. Informally $X_c$ is simply giving the distribution of the travel decisions of the users belonging to $c$. It is then natural to define the cumulated flows of the users of category $c$ following route $r$ as:

$$X_{rc}(I) := X_c(\{r\} \times I) \text{ for all } I \subseteq \mathcal{H} \tag{7.3}$$

**Remark 7.1.** *This very general representation of the user choices might seem more complicated than required. Indeed the distribution $D_c$ allows to represent non-discrete choice distributions for the users characterized by $(c, h_p)$. Informally in our approach for each $(c, h_p)$, we have a probability distribution representing the spread of the users $(c, h_p)$ over the possible departure times. In traditional transportation models, the problem is downsized to the much more specific case where* users with the same characteristics choose the same departure time. *Obviously this is very attractive from a computational perspective. However we will see later that on network with tolls this has some severe drawbacks, notably regarding the existence of an equilibrium. It is worth noting that this specific case can easily be embedded in our general approach by the introduction of the following concept.*

**Definition 7.2.** *A departure distribution $D_c$ is said to be symmetric with respect to the departure times if there exists a measurable function $H_c : \mathcal{H}_p \to \mathcal{H}$ such that:*

$$D_c(R \times \text{graph } H_c) = X_c^p(\mathcal{H}_p)$$

*$H_c$ is referred to as the symmetric reduction of the measure $D_c$.*

*This concept is initially due to Mas-Colell (1984).*

[1]The definition of a Marginal is given in Chapter 3 pp 111. If $M$ is a measure on a Cartesian product $A \times B$, then the marginal of $M$ on $A$ is the measure on $A$ such that $M_A(I) = M(I \times B)$ for each measurable subset $I \subseteq A$.

Users are assumed to be selfish cost minimizer agents and thus their travel decision is the solution of the mathematical program (7.2). For given travel time and toll vectors $(\boldsymbol{\tau}_{Rc}, \boldsymbol{p}_{Rc})$ the resulting distribution $D_c$ should lie in the set:

$$F_D^c(\boldsymbol{\tau}_{Rc}, \boldsymbol{p}_{Rc}) \equiv \Big\{(h, r, h_p) \in \mathcal{S} \times \mathcal{H}_p : g(h, \tau_{rc}(h), p_{rc}(h)|c, h_p) = \min_{(h,r)} g(h, \tau_r(h), p_r(h)|c, h_p)\Big\} \quad (7.4)$$

The set $F_D^c(\boldsymbol{\tau}_{Rc}, \boldsymbol{p}_{Rc})$ can be interpreted as the best response set and thus $F_D^c$ is the best response correspondence. For a given state of the supply $(\boldsymbol{\tau}_{Rc}, \boldsymbol{p}_{Rc})$, it gives all the possible demand states resulting from the optimization of the travellers.

## 1.2 Transport supply

**Arc travel time and cost model.** Each arc is composed of two parts. The first part is a free flow part where users drive at a speed depending on their vehicle class (and thus on their category). The second part is a queuing part where all users, whatever their vehicle class, wait to exit the arc according to a FIFO discipline. The physics of an arc $a$ is summarized by:

1. an exit capacity $k_a$,
2. and a vector of free flow travel time functions $\boldsymbol{\tau}_{0aC} := (\tau_{0ac})_{c \in C}$.

Both $k_a$ and $\boldsymbol{\tau}_{0aC}$ can be time varying. Moreover, it is assumed that for all $c$, the free flow travel time function $\tau_{0ac}$ is continuous, differentiable almost everywhere and such that $\dot{\tau}_{0ac} > -1$.

The travel time function of a given user category, *i.e.* $\tau_{ac}$, depends on the incoming flows of each user's categories. In other terms, it depends on the vector of cumulated flows $\boldsymbol{Y}_{aC}$. The computation of the travel time functions $\boldsymbol{\tau}_{ac}$ is essentially the same as in Section 3 of Chapter 4 apart from the computation of the cumulated volume at the entrance of the queue.

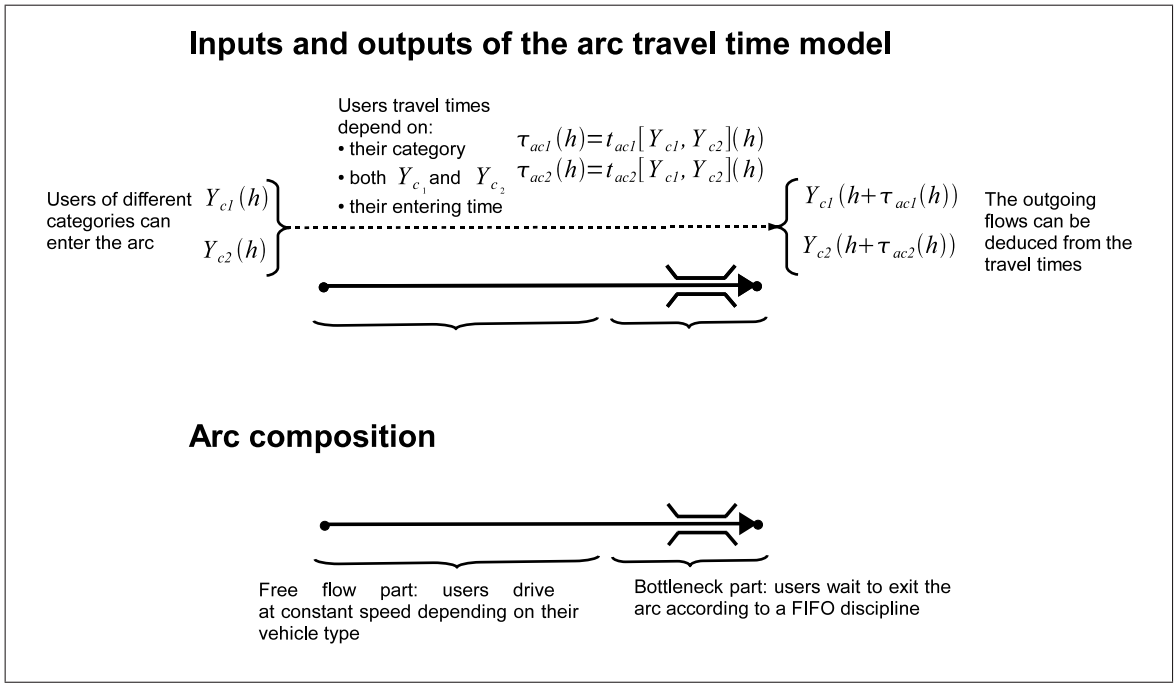



Figure 7.1: Arc travel time model principle

Let us now precise the arc travel time model. It is denoted $t_a$. As in previous Chapters, it is represented as a function on the space of cumulated flows $Y$, but returns a vector of arc travel time functions $(t_{ac}[Y])_{c\in C}$. , we can express the travel time model of an arc with the following system of equations. The stock of vehicle $Q$, seen as a function of the flow entering in the queue, denoted $\tilde{Y}$, arises from the equation:

$$\dot{Q}[\tilde{Y}](h) = \begin{cases} \tilde{y}(h) - k_a(h) & \text{if } Q[\tilde{Y}](h) > 0 \text{ or } \tilde{y}(h) > k \\ 0 & \text{otherwise} \end{cases} \tag{7.5}$$

The flow entering in the queue is $\tilde{Y} := \sum_{c\in C} Y_{ac} \circ H_{0ac}^{-1}$ where $H_{0ac} := \mathrm{id}_{\mathcal{H}} + \tau_{0ac}$. Then $\tau_{ac} = \tau_{ac}[\boldsymbol{Y}_{aC}]$ is the solution of the following equation:

$$K_a\left(h + t[\boldsymbol{Y}_{aC}](h)\right) - K_a \circ H_0(h) = Q[\tilde{Y}] \circ H_0(h) \tag{7.6}$$

where $K_a(h) = \int_{-\infty}^{h} k_a(u)du$. More details on travel time computation can be found in (Leurent, 2003*b*).

Each arc of the network is endowed with a toll function vector $\boldsymbol{p}_{aC} = (p_{ac})_{c\in C}$. Toll functions are functions of the time that indicates the monetary costs to cross an arc when entering at a given time. They are assumed to be continuous and differentiable almost everywhere.

**Route travel time and cost model.** When there is a single category of users, Chapter 3 formally defines the linkage between the route travel time functions and route vehicle flows. The problem of computing the route travel times is the Dynamic Network Loading problem (DNL) and an algorithm is presented in Appendix D.

Note the DNL can easily be extended to multi-category flows using the following procedure. Divide each arc in one arc for the queuing part and an arc by user category for the free flow part. Then adapt the flows to fit this new network. The new dynamic loading problem for multi-category flows is then equivalent to the one for single category flows. Figure 7.2 depicts this operation.

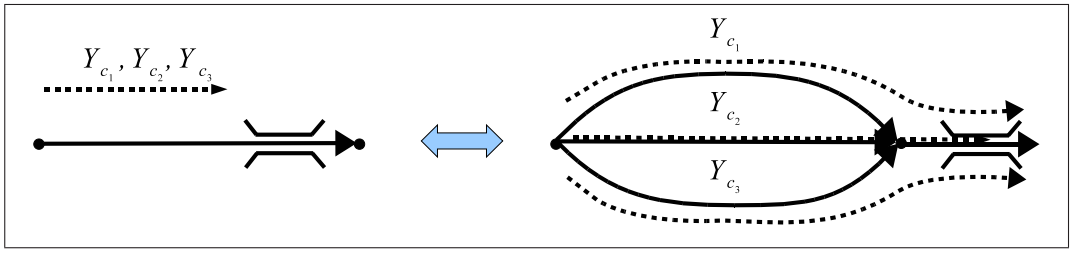



Figure 7.2: Multi-class dynamic loading principle
A simple example for one arc with three vehicle classes. The flow represented by $(Y_{c_1}, Y_{c_2}, Y_{c_3})$ is dispatched among three different routes on the new network.

The toll of a route is simply derived by summing the toll functions of each arc along the route, each of them being evaluated at the correct time of entrance.

For the sake of readability, we will adopt the compact notation $(\boldsymbol{\tau}_{RC}, \boldsymbol{p}_{RC}) = F_S(\boldsymbol{X}_{RC})$. An important property of the supply model is that the resulting route travel time functions $(\boldsymbol{\tau}_{RC}, \boldsymbol{p}_{RC})$ are continuous, FIFO and differentiable almost everywhere.

## 1.3 Equilibrium statement

In the previous subsections, a dynamic framework for transportation modelling has been set up. Supply is represented by a dynamic transport network $(A, N, \boldsymbol{t}_{AC}, \boldsymbol{p}_{AC})$. Each arc of the network endowed with a bottleneck model $t_a$ and toll functions $p_{ac}$. Demand is represented by a vector of preferred

arrival time distributions $\boldsymbol{X}_C^p := (X_c^p)_{c\in C}$. Let us now define precisely the notion of equilibrium in this context.

**Definition 7.3** (**D**ynamic **U**ser **E**quilibrium (DUE) )**.** *Find a departure distribution vector* $\boldsymbol{D}_C$ *such that for all* $c \in C$ :

$$D_c(B^c) = X_c^p(\mathcal{H}_p) \tag{7.7}$$

*with:*

$$B^c = F_D^c(\boldsymbol{\tau}_{Rc}, \boldsymbol{p}_{Rc}) \tag{7.8}$$

$$(\boldsymbol{\tau}_{RC}, \boldsymbol{p}_{RC}) = F_S(\boldsymbol{X}_{RC}) \tag{7.9}$$

Equation (7.9) encompasses the supply model while Equation (7.8) guarantees the optimality of the user travel decision and thus represents the demand model. Equation (7.7) simply states the balance between supply and demand. Note that apart from the fact that users are minimizing travel costs rather than maximizing utilities, this is the framework of dynamic congestion games as defined in Chapter 3.

# 2 A general property of the user equilibrium: the natural order of arrival

It has been shown in Chapter 5 that, on a single, untolled arc, the DUE exhibited a singular property. There always exists an equilibrium where users leave in the order of their arrival preferences. This so-called natural order of departure does not stand in the general case that has just been exposed.

Although the natural order of departure is no longer valid on a network with tolls, a similar property can be stated on the order of arrival. Before stating it, let us consider the problem from a new perspective. Consider a distribution $D_c$ representing an assignment of the demand and $(\boldsymbol{\tau}_{Rc}, \boldsymbol{p}_{Rc}) = F_S(\boldsymbol{X}_{RC})$ with $X_{rc}$ the marginal of $D_c$ on $\mathcal{S} \times \{r\}$. Then denote $H_{rc} := \mathrm{id}_{\mathcal{H}} + \tau_{rc}$ the route exit time function for category $c$. We define the *arrival distribution* $\bar{D}_c$ as:

$$\bar{D}_c(\{r\} \times H_{rc}(I) \times J) := \bar{D}_c(\{r\} \times I \times J) \tag{7.10}$$

$$\text{for all } r \text{ in } R,\ I \subset \mathcal{H} \text{ and } J \subset \mathcal{H}_p$$

$\bar{D}_c$ simply reinterprets $D_c$ by representing travel decisions under the form of a route and an arrival time rather than a route and a departure time. As we have defined symmetric departure distributions $D_c$, we can define symmetric arrival distributions $\bar{D}_c$.

**Definition 7.4.** *An arrival distribution $\bar{D}_c$ is said to be symmetric with respect to the arrival times if there exists a measurable function $\bar{H}_c : \mathcal{H}_p \to \mathbb{R}$ such that:*

$$\bar{D}_c(R \times \text{graph } \bar{H}_c) = X_c^p(\mathcal{H}_p)$$

*$\bar{H}_c$ is referred to as the symmetric reduction of the measure $\bar{D}_c$.*

This definition is the exact transposition of Definition 7.2 to arrival distributions. Redefining the users' travel decisions with respect to the arrival times rather than departure times might seem awkward in a model of departure time choice. Yet this point of view, although less natural, is extremely fruitful, as illustrated by the following results.

**Proposition 7.5.** *Assume given a state of the supply $(\boldsymbol{\tau}_{RC}, \boldsymbol{p}_{RC})$ and consider a user category $c$ with convex schedule delay cost function $D$ such that $D(0) = 0$. If two elements $(r_1, h_1, h_p^1)$ and $(r_2, h_2, h_p^2)$ of $\mathcal{S} \times \mathcal{H}_p$ are such that $(r_i, h_i)$ is the solution of the user optimization program for $(c, h_p^i)$, $h_p^1 \leq h_p^2$ and $h_1 + \tau_{r_1 c}(h_1) \geq h_2 + \tau_{r_2 c}(h_2)$ then:*

$$g(h_1, \tau_{r_1 c}(h_1), p_{r_1 c}(h_1) | c, h_p^i) = g(h_2, \tau_{r_2 c}(h_2), p_{r_2 c}(h_2) | c, h_p^i) \text{ for } i = 1 \text{ or } 2$$

Proposition 7.5 may seem technical, but it has a simple interpretation. Consider two users $(c, h_p^1)$ and $(c, h_p^2)$ such that $h_p^1 \geq h_p^2$ and assume they have chosen their arrival time in the reverse order of their arrival time preference (*i.e.* that they have chosen to arrive at $h_1, h_2 : h_1 \leq h_2$). Then Proposition 7.5 states they can switch their arrival decisions costlessly.

The proof of Proposition 7.5 is given in Appendix C.

To produce a general result on the order of arrival, let us introduce an additional assumption on the demand *i.e.* on $\boldsymbol{X}_C^p$. We say that $\boldsymbol{X}_C^p$ is atomless if for any $c$ the cumulative distribution $\boldsymbol{X}_c^p$ is continuous. A positive discontinuity (a cumulative distribution is increasing so it has no negative discontinuity) in $\boldsymbol{X}_c^p$ at an instant $h^p$ practically means that a large number of users have exactly the same arrival preferences. For instance, such a feature can be used to model the opening time of a factory.

Using Proposition 7.5, the following result can be established.

**Theorem 7.6** (On the order of arrival)**.** *Consider a DUE problem with atomless demand $\boldsymbol{X}^p_C$. Let $\boldsymbol{D}_C$ be a dynamic user equilibrium. Then there exists a dynamic user equilibrium $\boldsymbol{D}'_C$ such that the arrival distributions $(\bar{D}'_c)_{c\in C}$ are symmetric and that the symmetric reductions of $(\bar{D}'_c)_{c\in C}$ are non decreasing. Moreover for each category $c$ the marginal of $D'_c$ and $D_c$ on $\mathcal{S}$ are the equal.*

This theorem is especially interesting from a computational perspective. Indeed it states that when we are only interested in the flows on the networks (*i.e.* on the marginal of the equilibrium distributions), we can focus on symmetric distributions with respect to the arrival times. Now the set of symmetric distributions is much more easy to represent, as it allows to deal with functions, rather than distributions. In the following section, a reduced formulation of the DUE exploiting this result is presented.

The proof of Theorem 7.6 is technical in its details but simple in its principle. It uses Proposition 7.5 which states that given a DUE $\boldsymbol{D}_C$, one can rearrange all the users in their *natural order of arrival i.e.* such that if $(c, h^1_p)$ and $(c, h^2_p)$ are assigned respectively to $h_1$ and $h_2$ then $h^1_p > h^2_p \Leftrightarrow h_1 > h_2$. This concept is obviously very similar to the natural order of departure exposed in Chapter 5.

The fact that such a general model reveals such a strong property is particularly puzzling.

The proof of Theorem 7.6 is given in Appendix C.

# 3 A reduced formulation of the DUE problem

The previous section exposed an important property of the DUE problem. Whenever there exists a DUE, there also exists an equilibrium with a symmetric arrival distribution. Now the symmetric reduction of a distribution is much more easier to represent numerically so it is interesting to see how the dynamic user equilibrium definition can be restated in terms of symmetric arrival distributions.

## 3.1 The user's optimization program with arrival times

Let us first reformulate the program (7.2) with respect to arrival times rather than departure times. First note that route travel times and tolls can be eas-

ily reformulated as functions of the arrival time rather than of the departure time. For any route with travel time $\tau_{rc}$, this is achieved by composing $\tau_{rc}$ and $p_{rc}$ by the inverse of $H_{rc}$. We will denote them $\bar{t}_{rc}$ and $\bar{p}_{rc}$. For a given category $c = \{od, (D, \nu), u\}$, the new user's optimization program is

$$\min_{\bar{h},r} \nu \bar{t}_{rc}(\bar{h}) + \bar{p}_{rc}(\bar{h}) + D(\bar{h} - h_p)$$

Note that this program is equivalent to:

$$\min_{\bar{h}} \min_{r} \nu \bar{t}_{rc}(\bar{h}) + \bar{p}_{rc}(\bar{h}) + \min_{\bar{h}} + D(\bar{h} - h_p) \tag{7.11}$$

Equation (7.11) expresses that the user's optimization program can be decomposed in two steps: first find the optimal route $r$ for all optimal arrival times $\bar{h}$ and then find the optimal arrival time $\bar{h}^*$. In term of computing time, this is interesting as instead of scanning all the space $\mathcal{H} \times R$ to find a travel decision $(\bar{h}, r)$, it is possible to explore $\mathcal{H}$ and $R$ subsequently.

## 3.2 Representing travel decisions from an arrival time perspective

**Route choices.** At an aggregated level, the users route choices can be represented by functions $\bar{h} \mapsto \bar{R}_{rc}(\bar{h}, r)$ returning the proportions of users following a route $r$ and arriving at $\bar{h}$. This is the route choice function of category $c$. To be consistent with the optimal routes, the route choice function has to verify the following property:

$$\bar{R}_{rc}(\bar{h}) > 0 \Rightarrow r \text{ is an optimal route to arrive at } \bar{h} \text{ for users of category } c \tag{7.12}$$

**Arrival time choices.** To represent the arrival time choices, a natural approach is to introduce an arrival time choice function $\bar{H}_c$ that maps to each user of the category $c$ with preferred arrival time $h_p$ to his chosen arrival time $\bar{h} = \bar{H}_c(h)$. The function $\bar{H}_c$ will be assumed to be continuous and strictly increasing. Under this formalism, define the cumulative flow of a category $c$ at arrival as:

$$X_c^- := G_c \circ \bar{H}_c^{-1} \tag{7.13}$$

The demand $X_{rc}$ on each route is then obtained by a multiplication and a translation:

$$x_{rc} \circ H_{rc} := \bar{R}_{rc}.x_c^- \quad \text{with } H_{rc} := \mathrm{id}_{\mathcal{H}} + \tau_r \tag{7.14}$$

The operation of constructing the cumulated flows on each route from the route and arrival time choice functions of each user category $(\bar{\boldsymbol{R}}_{RC}, \bar{\boldsymbol{H}}_C)$ is denoted $F_D(\bar{\boldsymbol{R}}_{RC}, \bar{\boldsymbol{H}}_C)$.

## 3.3 Alternative formulations for the dynamic user equilibrium

Using the concepts introduced in the previous subsections, an alternative definition for the user equilibrium based on the variables $(\bar{H}_c, \bar{R}_c)$ rather than $D_c$ can easily be stated.

**Definition 7.7** (Dynamic user equilibrium based on arrival time functions)**.**

*Find* $(\bar{\boldsymbol{R}}_{RC}, \bar{\boldsymbol{H}}_C)$ *such that for every* $c = \big(od, (D, \nu), u\big)$ *and* $h_p$*:*

$$\bar{h}^\star = \bar{H}_c(h_p) \text{ and } \bar{R}_c(r, \bar{h}) > 0$$

$$\Rightarrow \nu.\bar{t}_{rc}(\bar{h}^\star) + p_{rc}(\bar{h}^\star) + D(\bar{h}^\star - h_p) = \min_{\bar{h}, r'} \ \nu.\bar{t}_{r'c}(\bar{h}) + p_{r'c}(\bar{h}) + D(\bar{h} - h_p) \quad (7.15)$$

*with:*

$$(\bar{\boldsymbol{\tau}}_{RC}, \bar{\boldsymbol{p}}_{RC}) = F_S(\boldsymbol{X}_{\boldsymbol{RC}}) \quad (7.16)$$

$$\boldsymbol{X}_{RC} = F_D(\bar{\boldsymbol{R}}_{RC}, \bar{\boldsymbol{H}}_C) \quad (7.17)$$

This formulation is much more suited for computational purposes than the original definition adopted in the first section of the chapter. As shown in Theorem 7.6 this is a natural approach, as a solution of the alternative formulation leads to an equilibrium in the sense of Definition 7.3. Reversely as soon as an equilibrium departure distribution $\boldsymbol{D}_C$ has been found, a solution $(\bar{\boldsymbol{R}}_{RC}, \bar{\boldsymbol{H}}_C)$ of the problem exposed in Definition 7.7 can be computed. Note that in Definition 7.7 the unknown variables are the arrival time and the route functions. It is might be more convenient to work directly with the cumulated volumes. Hence the following definition.

**Definition 7.8** (Dynamic user equilibrium based on cumulated volumes)**.**
*Find* $\boldsymbol{X}_{RC}$ *such that* $(\bar{\boldsymbol{R}}_{RC}, \bar{\boldsymbol{H}}_C)$ *is a solution to the DUE as stated in Definition 7.7, letting for all* $r, c$*:*

$$(\bar{\boldsymbol{\tau}}_{RC}, \bar{\boldsymbol{p}}_{RC}) = F_S(\boldsymbol{X}_{\boldsymbol{RC}}) \quad (7.18)$$

$$\bar{H}_c := (\sum_r \bar{X}_{rc})^{-1} \circ X_c^p \tag{7.19}$$

$$\bar{R}_{rc} := \bar{x}_{rc} / \sum_{r'} \bar{x}_{r'c} \tag{7.20}$$

$$\bar{X}_{rc} := X_{rc} \circ (\mathrm{id}_{\mathcal{H}} + \tau_{rc}) \tag{7.21}$$

The difference between Definitions 7.7 and 7.8 is merely a question of notations.

## 3.4 Measuring the quality of a solution

The analytical resolution of the DUE problem on a general network is hard. Instead, approximate methods are used. This requires a way to state if a candidate equilibrium is acceptable or not. Fortunately having a precise formulation at our disposal allows us to establish rigorous criterion to measure the quality of candidate solutions.

**Definition 7.9** (Least cost criterion)**.** *Consider route flow vector $\boldsymbol{X}_{RC}$ and define $\bar{H}_c$ and $\bar{X}_{rc}$ for all $r, c$ as in Definition 7.8. The least cost criterion $I(\boldsymbol{X}_{RC})$ is then:*

$$I(\boldsymbol{X}_{RC}) := \sum_{c \in C} \frac{1}{N_c} \sum_{r \in R} \int_{h_p \in \mathcal{H}_p} \frac{\tilde{g}_{rc}(h_p) - g_c^\star(h_p)}{\tilde{g}_{rc}(h_p)} dX_c^p(hp) \tag{7.22}$$

*with for all $r, c$:*

$$\begin{aligned} N_c &:= X_c^p(\mathcal{H}_p) \\ \tilde{g}_{rc}(h_p) &:= \nu \bar{t}_{rc}\big(\bar{H}_c(h_p)\big) + \bar{p}_{rc}(\bar{H}_c(h_p)) + D\big(\bar{H}_c(h_p) - h_p\big) \\ g_c^\star(h_p) &:= \min_{r,\bar{h}} \nu \bar{t}_{rc}(\bar{h}) + \bar{p}_{rc}(\bar{h}) + D(\bar{h} - h_p) \end{aligned}$$

In Definition 7.9, the function $\tilde{g}_{rc}$ can be interpreted as the cost incurred by a user of category $c$ and preferred arrival timer $h_p$ following the route $r$. The function $g_c^\star$ gives the minimal cost for a user characterized by $(c, h_p)$. This gives a simple economic interpretation to the least cost criterion: it is the average cost that users could save by rerouting and rescheduling their trip. Naturally $I(\boldsymbol{X}_{RC}) = 0$ if and only if $\boldsymbol{X}_{RC}$ is an equilibrium.

## Conclusion

This chapter explores various formulations for the dynamic user equilibrium with departure time choice on a tolled network. At first a general formulation inspired by the framework of dynamic congestion games is proposed. In this approach users with the same characteristics can choose different departure times. Then it is shown that it is possible to search only for equilibriums with symmetric arrival distributions (*i.e.* where users with the same characteristics always choose the same arrival time) without compromising the existence of an equilibrium. Finally two reduced formulations for symmetric equilibriums are proposed. The first one is based on arrival time and route choice functions whereas the second one is based on route flows.

Chapter 8

# A convex combination algorithm to compute the dynamic user equilibrium

In this chapter a numerical scheme is proposed to compute the restricted formulation of the dynamic user equilibrium presented in Chapter 7. It was initially designed to extend the LADTA model introduced by Leurent (2003*b*) whose original implementation did not account for departure time choice. As we had at our disposal the LADTA toolkit, a powerful implementation of the main procedures of LADTA, one of our goals was to keep as much as possible the same philosophy as the original computation algorithm. The principle of LADTA is to consider the DUE computation as a fixed point problem, and to compute by a convex combination procedure.

In a nutshell the procedure is the following. Initialize the state of the network by assigning null traffic flows to the arcs and setting the travel times their free flow values. Then repeat iteratively the following process: (1) compute the least cost routes for each $OD$ pairs and assign the flows of traffic accordingly; (2) load the traffic on the arcs of the network and update the travel times (3) compute a convex combination of the resulting arc flows with the previous ones and store the results.

The most natural approach to extend this algorithm to incorporate a departure time choice model is to redesign step (1) in order to assign user not only according to the optimal route but according to the optimal transport service *i.e.* to the optimal pair of route and departure time. In this perspective extending LADTA to incorporate departure time choice is essentially being able to solve efficiently the users' minimization program.

This chapter is divided into four parts. First, a general overview of the

algorithm is presented (Section 1). Second, the two main steps, the arrival time assignment and the spreading procedure are presented in details (Sections 2 and 3). Finally, the algorithm performance is assessed in numerical examples on different networks (Section 4).

# 1 General presentation

## 1.1 Context of application

The notations of the previous chapter are used.

For computational reasons we restrict ourselves to piecewise linear functions (PWL functions). A PWL function will be encoded by a list of triple $(x_i, y_i, s_i)$. A triple is said to be a piece. $x$ belongs to the ith piece if $x \in [x_i, x_{i+1}]$.

In particular $X_c^p$, $p_{rc}$, $\tau_{0ac}$, $D$ and $K_a$ are PWL functions. As a consequence, equilibrium route flows $X_{rc}$ are also PWL. The resulting travel times functions $\tau_{rc}$ are then PWL from the properties of the bottleneck model.

## 1.2 LADTA solution method overview

LADTA is a model proposed for dynamic user equilibrium in (Leurent, 2003*b*). The physical and economic assumptions are the same as the one retained in our model except no departure time choice model exists in the current version of LADTA. In LADTA demand is described by a $OD$ matrix $\boldsymbol{X}_C := (X_c)_{c \in C}$ where $C$ is the set of user categories.

The solution method of LADTA is based on four procedures:

1. *The loading procedure*, to obtain the flows on each arc. It loads route traffic flows using a travel time function vector $\boldsymbol{t}_{AC}$. It is denoted $\boldsymbol{Y}_{AC} = F_L(\boldsymbol{X}_{RC}, \boldsymbol{t}_{AC})$.

2. *The traffic flowing procedure*, that computes the actual travel times on each arc from the arc inflow. It is denoted $\boldsymbol{t}_{AC} = F_F(\boldsymbol{Y}_{AC})$.

3. *The formation of services*, that computes the least cost route based on the arc travel time and toll functions. It is computed for each user's category *i.e.* on each origin destination pair and for all possible values of time and departure instants. It is a classical operational research

problem (see Chapter 1 for a review). The procedure stores the results in a least cost route tree denoted $\boldsymbol{R}_{AC}$. The procedure is denoted $\boldsymbol{R}_{AC} = F_{LC}(\boldsymbol{t}_{AC}, \boldsymbol{p}_{AC})$.

4. *The user's choice*, that computes route flows by assigning the OD matrix flows ($\boldsymbol{X}_C$) to the optimal route. It is denoted $\boldsymbol{X}_{RC} = F_D(\boldsymbol{R}_{AC}, \boldsymbol{X}_C)$.

For more information on each of the procedures see (Leurent, 2003*b*). Some more details are also available in Chapter 2, section 3, page 94. The algorithm then consists of iteratively applying these four procedures in a convex combination scheme. The algorithm is made explicit below (Algorithm 8.1).

**Algorithm 8.1** `LADTA RouteChoice`($\boldsymbol{X}_C$)

**Inputs:** An $OD$ matrix $\boldsymbol{X}_C$
**Outputs:** $\boldsymbol{Y}_{AC}$ the arc cumulated flows for each user category
**Parameter:** $w_k$ a decreasing sequence from 1 to 0
**Initialize** $\boldsymbol{Y}_{AC}^{[0]} := 0$ and $k := 0$
**Repeat**
**Set** $\boldsymbol{t}_{AC} := F_F(\boldsymbol{Y}_{AC}^{[k-1]})$
**Set** $\boldsymbol{R}_{AC} := F_{LC}(\boldsymbol{t}_{AC}, \boldsymbol{p}_{AC})$
**Set** $\boldsymbol{X}_{RC} := F_D(\boldsymbol{R}_{AC}, \boldsymbol{X}_C)$
**Set** $\boldsymbol{Z}_{AC} := F_L(\boldsymbol{X}_{RC}, \boldsymbol{t}_{AC})$
**Set** $\boldsymbol{Y}_{AC}^{[k]} := w_k.\boldsymbol{Y}_{AC}^{[k-1]} + (1 - w_k).\boldsymbol{Z}_{AC}$
**Set** $k := k + 1$
**Until** $\boldsymbol{Y}_{AC}^{[k]}$ satisfies a certain criterion.
**End For**

## 1.3 Philosophy of the algorithm for combined route and departure choice

As mentioned in the introduction, a guideline in the design of the algorithm was to make the most of the LADTA ToolKit (LTK). That's why for the supply side, we (purposely) adopted the same modelling choices. Essentially, only the user choice procedure ($F_D$) and the demand description ($\boldsymbol{X}_C$) have to be changed.

The demand is now an OD matrix in preferred arrival times and not in actual times as before. The quantity $\boldsymbol{X}_C^p = (X_c^p)_{c\in C}$ represents the OD matrix: each element *od* of the *OD* matrix is a sequence of distributions $(X_c^p)_{c:od\in c}$ that can be interpreted as cumulated flows. The user decision procedure was assigning flows to the optimal route. It will now assign the flows to the optimal route and departure time.

To state the new algorithm, the previous procedures need to be slightly redefined. Except from the user's choice procedure, they need to be expressed with variables from an arrival time perspective. Overall the changes are minor.

1 The loading procedure now loads route flows at arrival on the network. It propagates backward the arrival flows, using a travel time function vector $\boldsymbol{t}_{AC}$. It is denoted $\boldsymbol{Y}_{AC} = F_L(\bar{\boldsymbol{X}}_{RC}, \boldsymbol{t}_{AC})$.

2 The traffic flowing procedure is the same. It is still denoted $\boldsymbol{Y}_{AC} = F_F(\boldsymbol{X}_{RC}, \boldsymbol{t}_{AC})$

3 The formation of services, now computes least cost routes to arrive at a given arrival instant. The review in Chapter 1 presents efficient algorithms for this procedures. In addition to the least cost route tree, it now returns a vector of least cost functions $\bar{\boldsymbol{g}}_C = (g_c)_{c\in C}$. For a given category $c$, the least cost function maps an arrival time with the least cost to arrive at this time. The procedure is now is denoted $\bar{\boldsymbol{R}}_{AC}, \bar{\boldsymbol{g}}_C = F_{LC}(\boldsymbol{t}_{AC}, \boldsymbol{p}_{AC})$.

4 The user's choice now computes route flows by assigning the OD matrix flows $\boldsymbol{X}_C$ to the optimal route and arrival times. It is denoted $\boldsymbol{X}_{RC} = F_D(\bar{\boldsymbol{R}}_{AC}, \bar{\boldsymbol{g}}_C, \boldsymbol{X}_C^P)$.

The pseudo-code of the new algorithm is now:

**Algorithm 8.2** `LADTA RouteDepartureChoice`$(\boldsymbol{X}_C^p)$

**Inputs:** An $OD$ matrix in preferred arrival times $\boldsymbol{X}_C^p$
**Outputs:** $\boldsymbol{Y}_{AC}$ the arc cumulated flows for each user category
**Parameter:** $w_k$ a decreasing sequence from 1 to 0
  **Initialize** $\boldsymbol{Y}_{AC}^{[0]} := 0$ and $k := 0$
  **Repeat**
    **Set** $\boldsymbol{t}_{AC} := F_F(\boldsymbol{Y}_{AC}^{[k-1]})$
    **Set** $\bar{\boldsymbol{R}}_{AC}, \bar{\boldsymbol{g}}_C := F_{LC}(\boldsymbol{t}_{AC}, \boldsymbol{p}_{AC})$
    **Set** $\bar{\boldsymbol{X}}_{RC} := F_D(\bar{\boldsymbol{R}}_{AC}, \bar{\boldsymbol{g}}_C, \boldsymbol{X}_C)$
    **Set** $\boldsymbol{Z}_{AC} := F_L(\bar{\boldsymbol{X}}_{RC}, \boldsymbol{t}_{AC})$
    **Set** $\boldsymbol{Y}_{AC}^{[k]} := w_k.\boldsymbol{Y}_{AC}^{[k-1]} + (1 - w_k).\boldsymbol{Z}_{AC}$
    **Set** $k := k + 1$
  **Until** $\boldsymbol{Y}_{AC}^{[k]}$ satisfies a certain criterion.
  **End For**

The user's decision still need to be precisely defined. The global philosophy of the user's decision problem have already been stressed out by the second formulation of the DUE problem (see previous Chapter), whose primary aim was to ease the algorithmic. Figure 8.1 exposes how to compute route flows from the $OD$ matrix and travel time and toll functions on the arcs. It can be summarized as follows. By computing the optimal arrival times for each user, one can deduce the traffic volumes at arrival. Then, by computing the optimal route in order to arrive at a destination $d$ for every instant $h$, one can obtain in turn the route flows at arrival. However we will see later that directly computing the arrival flows from the optimal arrival times would lead to discontinuous cumulated flows on the arcs. As this is inconvenient from a computational viewpoint, a spreading procedure is proposed.

Consequently the user's choice procedure is divided in three sub-procedures:

- The arrival time choice procedure that computes a PWL function vector $\bar{\boldsymbol{H}}_C := (\bar{H}_c)_{c \in C}$ on the basis of the least cots routes. For a given category $c$, the function $\bar{H}_c$ maps a preferred arrival time with the corresponding optimal arrival time.

- The spreading procedure computes the flows at arrival from $\bar{\boldsymbol{H}}_C$ and the least cost routes. The flows at arrival are denoted $\bar{\boldsymbol{X}}_C$.
- The route choice procedure simply splits the flows $\bar{\boldsymbol{X}}_C$ among the routes. For each category, it assigns the cumulated flows $\bar{X}_c$ on the corresponding least cost routes. It returns a route flow vector $\bar{\boldsymbol{X}}_{RC}$ expressed as a function of the arrival times.

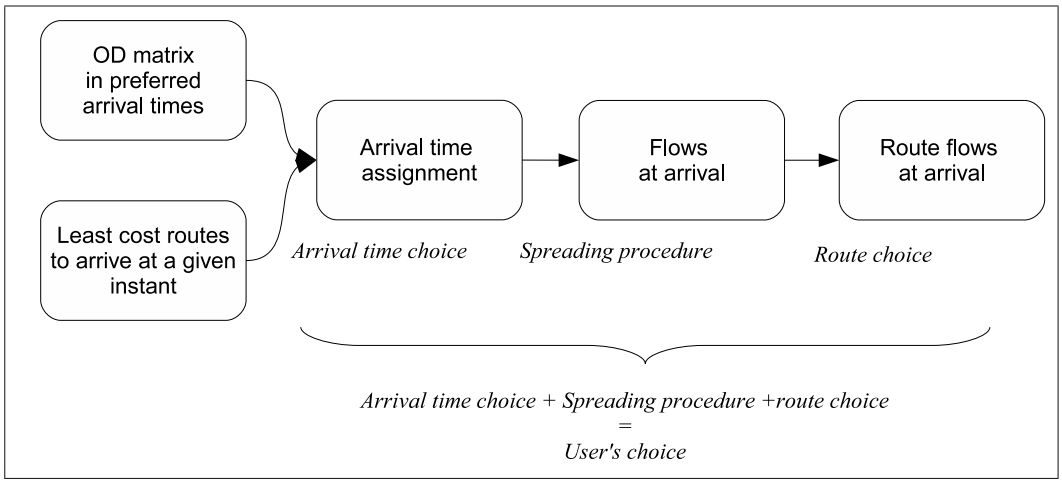



Figure 8.1: Chart of the route flows computation

To complete the precise description of algorithm 8.2, the user's choice procedure needs to be precisely defined. This is achieved in the two subsequent sections. The first one (Section 2) is dedicated to the arrival time choice, while the second one (Section 3) describes the spreading procedure.

## 2 Optimal trip scheduling

In this section, an exact algorithm to compute the optimal arrival time for all users of a given category $c$ is presented. The idea is to treat conjointly all the users in an event based approach. For the sake of clarity, it is first assumed that the users' schedule delay cost function has the V-shaped form: $D(l) := \alpha l^+ + \beta l^-$. The last subsection explains how to alleviate this assumption.

## 2.1 Algorithm statement

Let the travel time and toll functions $\bar{t}_{rc}$ and $\bar{p}_{rc}$ be given for all the routes of the network and consider a specific user category $c = (od, (\nu, D), u)$. The functions $(\bar{t}_{rc}, \bar{p}_{rc})_{r\in R}$ are expressed as functions of the arrival time. Define

$$g_c(\bar{h}) := \min_{r\in R_{od}} \nu \bar{t}_{rc}(\bar{h}) + \bar{p}_{rc}(\bar{h})$$

In this subsection the problem of computing the function $\bar{H}_c$, described by the following equation, is addressed.

$$\begin{aligned} \bar{H}_c(h_p) = \min \big\{ & \bar{h}^\star \text{ such that:} \\ g_c(\bar{h}^\star) + D(h_p - \bar{h}^\star) = & \min_{\bar{h}\in\mathcal{H}} g_c(\bar{h}) + D(h_p - \bar{h}) \big\} \end{aligned} \tag{8.1}$$

This problem boils down to the resolution of a continuum of optimization programs, one for each $h_p \in \mathcal{H}_p$. One might be tempted to discretize $\mathcal{H}_p$ and then to solve distinctly the resulting problems. However it is natural to think that the program corresponding to $h_p$ has something to do with the program for $h_p + dh$. The global idea of this algorithm is to work in this direction.

First consider a given $h_p$. Let us write the first order condition. As the functions $\bar{t}_{rc}$, $\bar{p}_{rc}$ and $D$ are not differentiable everywhere on $\mathcal{H}$, the concept of subdifferential is used[1].

$$0 \in \partial(g_c - D)(\hbar) \tag{8.2}$$

$\partial D(\bar{h})$ can be either $\{\alpha\}$, $\{\beta\}$ or $[\alpha, \beta]$. This leads to three cases to consider:

1. For $\bar{h} < h_p$, only the arrival times such that $\alpha \in \partial g_c$ needs to be considered.
2. For $\bar{h} > h_p$, only the arrival times such that $\beta \in -\partial\big\{-g_c\big\}$ needs to be considered.
3. For $\bar{h} = h_p$, the first order condition is met if and only if $\partial g_c \cap [\alpha; \beta] \neq \emptyset$.

[1] In real analysis, the subdifferential of $f$ in $x$ is the set $[\lim_{x\to a^-} f; \lim_{x\to a^+} f]$ (or $\emptyset$ if it does not make sense) and is denoted $\partial f(a)$. Then $0 \in \partial f(a)$ is a necessary condition for $f$ to admit an extrema in $a$.

That is to say that we can group candidate optimal arrival times in three categories: early candidates $\bar{h}_i^e$, late candidates $\bar{h}_i^l$ and possibly the preferred arrival time. These three cases are depicted in Figure 8.2.

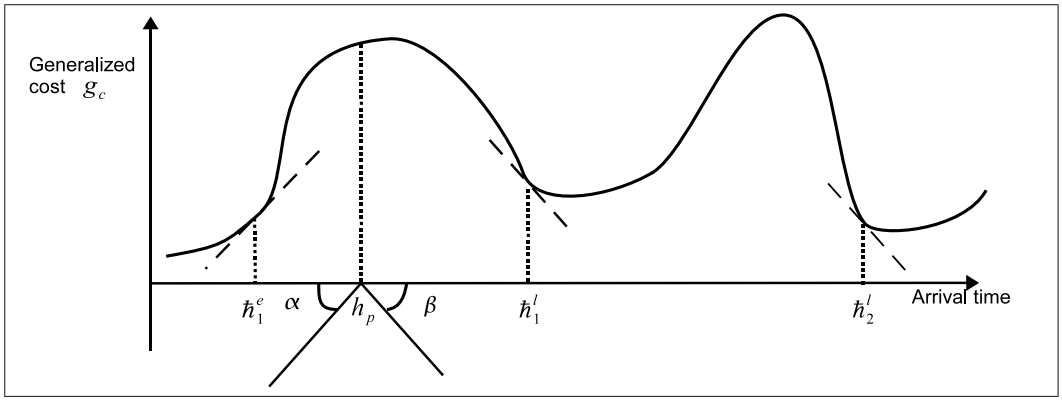



Figure 8.2: The first order condition

When the functions considered are limited to continuous PWL functions, computing those three groups is straightforward. Indeed $\partial g_c(x)$ is easy to compute. If $x$ belongs to a single piece $i$ then $\partial g_c(x) = \{s_i\}$. Otherwise it belongs to two successive pieces, say $i$ and $i+1$, and $\partial g_c(x) = [s_i, s_{i+1}]$ (or $\emptyset$ if it does not make sense). A simple scan of the list of the pieces is hence enough to compute the candidate arrival times $\bar{h}_i^e$ and $\bar{h}_i^l$.

Algorithm 8.3 explicits in pseudo code the method `FindTangencyPoint`$(g_c, z)$. It finds for any PWL function $g_c$ and real $\alpha$, the list of points $(h_i)$ such that $z \in \partial g_c(h_i)$. Applying `FindTangencyPoint` on $(g_c, \alpha)$ and on $(-g_c, \beta)$ then allows to compute the lists $(\bar{h}_i^e)$ and $(\bar{h}_i^l)$.

Now let us return to the original problem. Once the optimal departure time $\bar{h}$ for a preferred arrival time $h_p$ has been computed, how can we deduce the optimal $\bar{h}'$ for $h_p + dh$? The answer arises from two remarks. First, if $dh$ is sufficiently small, the candidate arrival times resulting from `FindCandidate` are roughly the same (see Figure 8.2). The only possible changes are: the withdrawal of $\bar{h}_1^l$ because $\bar{h}_1^l < h_p + dh$ or the addition of an early instant $\bar{h}_1^e$. Second, the variation between $g_c(\bar{h}) + D(\bar{h} - h_p)$ and $g_c(\bar{h}) + D(\bar{h} - h_p + dh)$ is straightforward to establish for any $\bar{h}$. Thus while varying $h_p$ one can easily track the evolution of the generalized cost for early candidate instants (*i.e.* $g_c(\bar{h}_i^e) + D(\bar{h}_i^e - h_p)$), for late ones (*i.e.* $g_c(\bar{h}_i^l) + D(\bar{h}_i^l - h_p)$) as well as the

generalized cost for the preferred arrival time (*i.e.* $g_c(h_p)$). The only point to be careful about is to update the two lists of candidate arrival times when necessary.

This gives the general lines of the master procedure (Algorithm 8.5). First we find the optimal $\bar{h}$ for $h_p^{\min} = \min\ \mathcal{H}_p$ the function $\bar{H}$ is initialized with $(h_p^{\min}, \bar{h}, 1)$ if $\bar{h} = h_p$ and $(h_p^{\min}, \bar{h}, 0)$ otherwise. Then the next "event" is a change in the slope (if $\bar{h} = h_p$) or in the optimal arrival time. The pseudo-code of the sub algorithm 8.4 details how to compute the next event. For the event of type (1) $\bar{H}$ is updated, while for event of type (2) the list of candidate arrival times is. The process is iterated until all $\mathcal{H}_p$ has been covered. The iteration is described in Algorithm 8.5.

**Algorithm 8.3** `FindTangencyPoint`$(g, z)$

**Inputs:** A function $g$ and a real $z$
**Outputs:** A list of real $(\bar{h}_i)$
**Set** $s_{\text{previous}} \leftarrow 0$
**Foreach** pieces $(x_i, y_i, s_i)$ of $g$
**if** $z \in [s_{\text{previous}}; s_i]$ **then** Add $x_i$ to the list $(\bar{h}_i)$
**End For**

**Algorithm 8.4** `FindNextEvent`$(g, h_p, (h_i^e), (h_i^l), \alpha, \beta)$

**Inputs:** A PWL function $g$, three positive reals and two lists
**Outputs:** A triple $(\tilde{h}_p, \bar{h}, s)$
**Set** $h_m^e = \arg\min_i g(h_i^e) + \alpha(h_p - h_i^e) : h_i^e < h_p$
and $h_m^l = \arg\min_i g(h_i^l) + \beta(h_i^l - h_p) : h_i^l > h_p$
**Switch**
**Case:** $g(h_m^e) = \min\{g(h_m^e), g(h_m^l), g(h_p)\}$
**Solve** $g(\bar{h}) = g(h_m^e) + \alpha.\bar{h}$ and $g(h_m^e) + \alpha.\bar{h} = g(h_l^e) - \beta.v$ for$\bar{h} > h_p$
**Set** $\bar{h}$ to the minimum of the two solutions and $\tilde{h}_p$ and $s$ to respectively $h_p$ and 1 or $h_m^l$ and 0 accordingly.
**Case:** $g(h_l^e) = \min\{g(h_m^e), g(h_m^l), g(h_p)\}$
**Solve** $g(\bar{h}) = g(h_m^e) - \beta\bar{h}$ and $g(h_m^e) + \alpha\bar{h} = g(h_m^l) - \beta\bar{h}$ for$\bar{h} > h_p$
**Set** $\bar{h}$ to the minimum of the two solutions and $\tilde{h}_p$ and $s$ to respectively $h_p$ and 1 or $h_m^e$ and 0 accordingly.
**Case:** $g(h_p) = \min\{g(h_m^e), g(h_m^l), g(h_p)\}$
**Solve** $g(\bar{h}) = g(h_m^l) - \beta\bar{h}$ and $g(\bar{h}) = g(h_m^e) + \alpha\bar{h}$ for$\bar{h} > h_p$
**Set** $\bar{h}$ to the minimum of the two solutions and $\tilde{h}_p$ to $h_m^l$ or $h_m^e$ accordingly and $s$ to 0
**End Switch**

**Algorithm 8.5** `FindOptArrivalTime`$(g, \alpha, \beta)$

**Inputs:** A PWL function $g$ and two positive reals
**Outputs:** A PWL function $H$

**Set** $h_p$ to $\min \mathcal{H}_p$
**Set**$(h_i^e)$ to `FindTangencyPoint`$(g, \alpha)$
**Set** $(h_i^l)$ to `FindTangencyPoint`$(g, \beta)$
**While** $h_p < \max \mathcal{H}_p$
**Set** $(h_p, \bar{h}, s) \leftarrow$`FindNextEvent`$(g, h_p, (h_i^e), (h_i^l), \alpha, \beta)$
**If** $\bar{h} < \min\{h_i^l : h_i^l > h_p\}$ then **add** $(h_p, \bar{h}, s)$ to $H$
**set** $h_p$ to $\min h_i^l : \{h_i^l : h_i^l > h_p\}$
else **set** $h_p$ to $\bar{h}$
**End While**

**Remark 8.1.** *In Algorithm 8.4, the computation of* $\arg\min_i g_c(h_i^e) + \alpha(h_p - h_i^e) : h_i^e < h_p$ *and* $\arg\min_i g(h_i^l) + \beta(h_i^l - h_p) : h_i^l > h_p$ *can be optimized by keeping in memory the values* $g_c(h_i^e)$ *and* $g_c(h_i^l)$ *while executing Algorithm 8.3.*

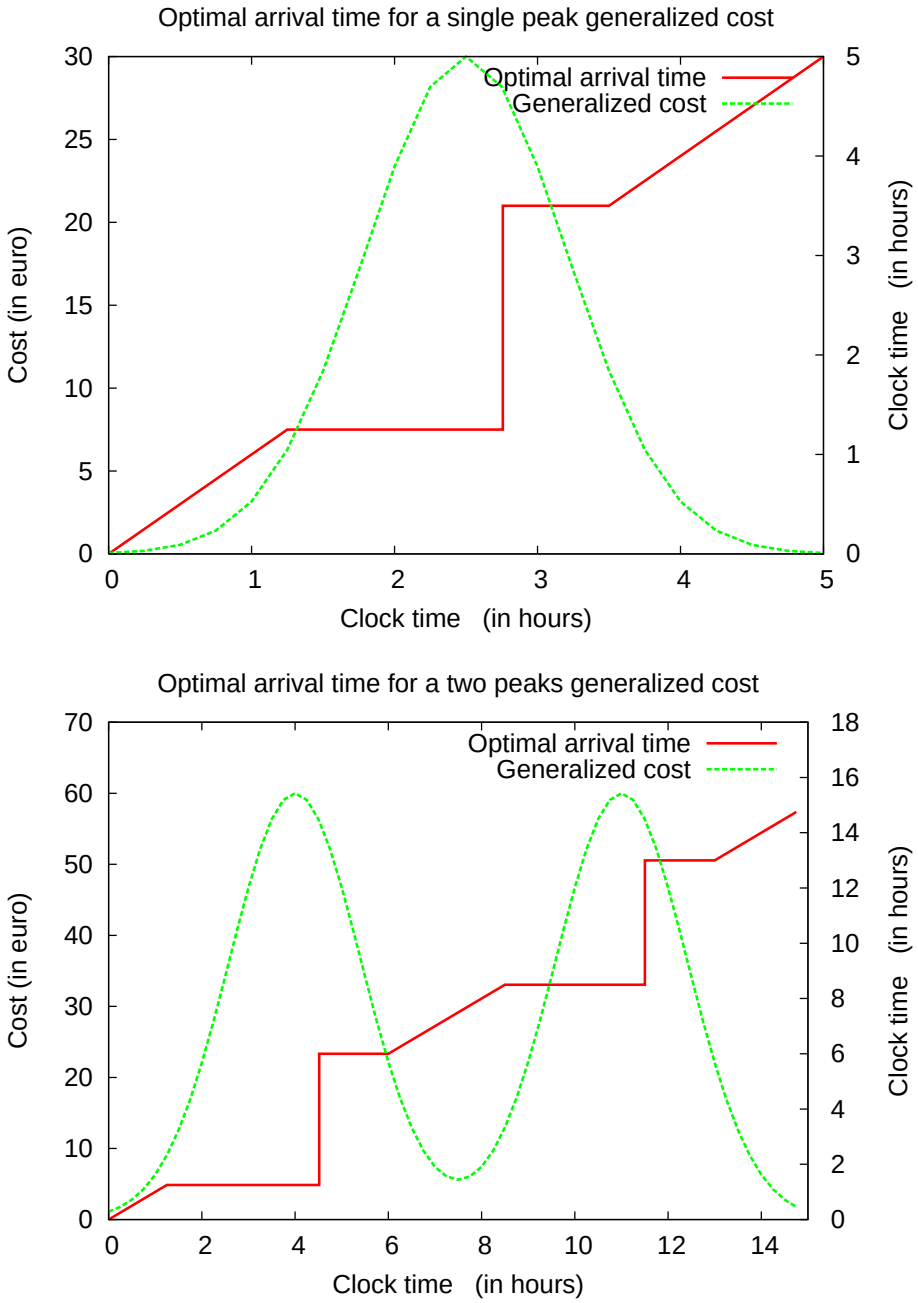



Figure 8.3: Two simple numerical illustrations

## 2.2 Numerical illustrations, complexity analysis and benchmark

**Numerical illustrations.** Figure 8.3 gives two simple illustrations of optimal arrival time functions. In the first example the generalized cost at arrival has the shape of a Gaussian centered in $\bar{h} = 5$ and hence admits a unique maximum. The optimal arrival time function for $\alpha = 0.5$ and $\beta = 1.5$ is plotted on the same figure. When users have preferred times close enough to the peak in generalized cost, their optimal arrival time is either delayed after or before the peak according to the relative value of $\alpha$ and $\beta$. In the second example $g_c$ admits two maxima. The pattern of $\bar{H}$ is quite similar: at the beginning or at the end of the period, as well as between the two peaks, users choose to arrive at their preferred time. On the contrary they reconsider their arrival time in peak periods.

Figure 8.4 shows two examples of the results generated with more complex generalized cost functions obtained by summing 50 Gaussian functions with random mean. Note that it is still easy to interpret $\bar{H}_c$ by examining the multiple peaks in $g_c$.

**Complexity analysis.** Denote $n$ the number of pieces in $g_c$ and $p$ the number of the function returned by the procedure `FindOptArrivalTime`. In Algorithm 8.5 the running time can be divided in two parts. The initializations of the tangency points $(h_i^e)$ and $(h_i^l)$ requires a full scan of $g_c$ which can be achieved in $O(n)$. The main loop running time is the number of events treated (i.e. approximatively $p$) times the amount of time required to compute an event. In Algorithm 8.4 the time-consuming operations are the resolution of the equations involving $g_c$; they can be solved using a simple scan forward which lasts approximately $n/p$. Consequently the main loop has a complexity of $O(n + p)$. The following proposition comes:

**Proposition 8.2.** *The running time of* $\bar{H}$ =`FindOptArrivalTime`$(g, \alpha, \beta)$ *is* $O(n+p)$*, where* $n$ *is the number of pieces in* $g_c$ *and* $p$ *the number of pieces in the resulting PWL function* $\bar{H}$*.*

Obviously this proposition only partially answers the question of the complexity of our algorithm as the relation between $p$ and $g_c$ is not established. To the author's opinion this is *a priori* a difficult question, as there are

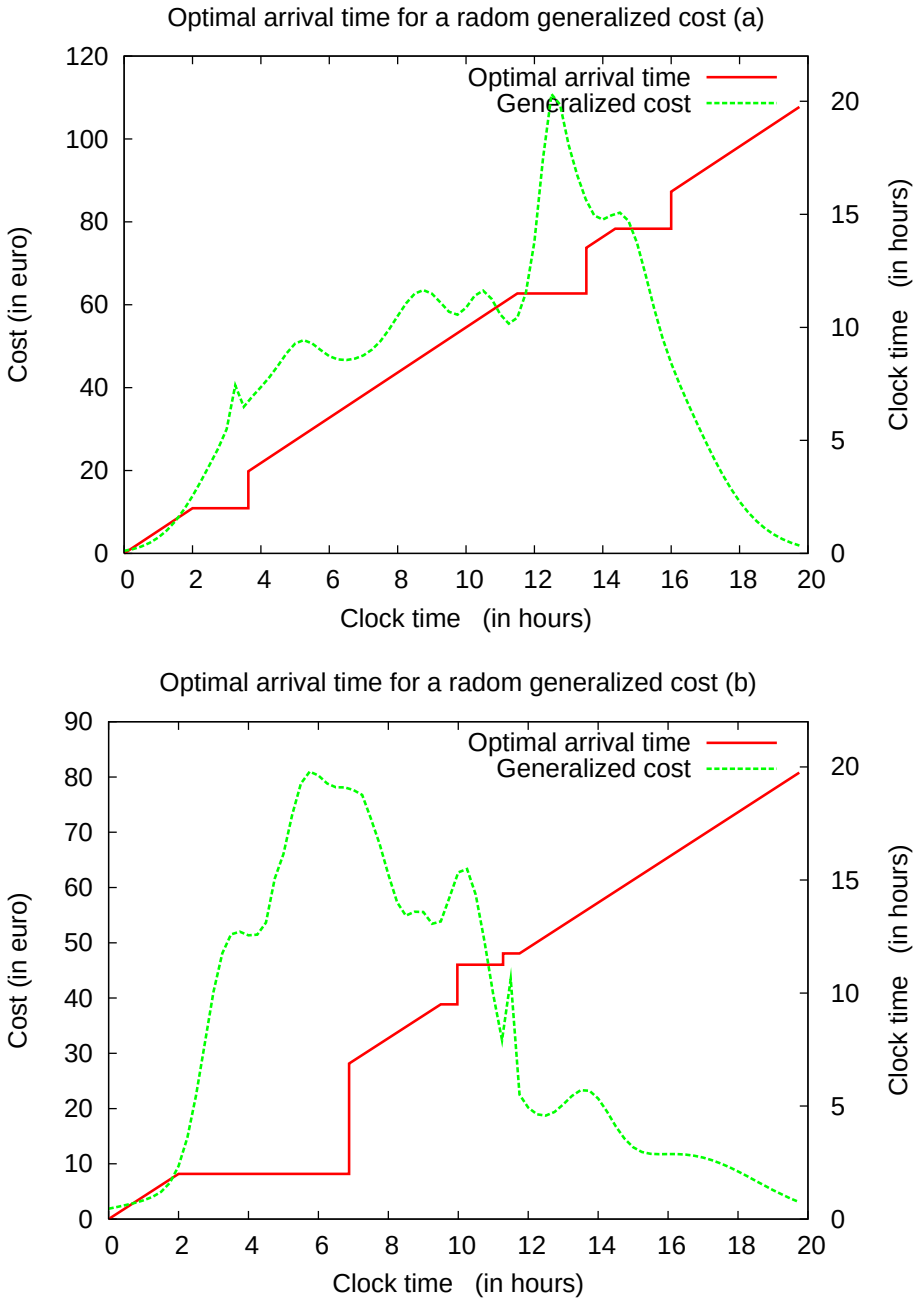



Figure 8.4: Two complex illustrations

seemingly no other options than actually computing $H$ from $g_c$ to establish $p$. Hence a numerical investigation is conducted below.

Second it is interesting to compare this result to the complexity of the naïve computing procedure consisting in discretizing $\mathcal{H}_p$ in $k$ reals and solving independently the corresponding sequence of minimization programs. Each minimization program has a complexity of $O(n)$ so the global running time is in $O(kn)$. The question arising is how $k$ should be chosen to give an acceptable approximation of $\bar{H}$. It is quite clear that more pieces $\bar{H}$ has, the higher $k$ should be. A reasonable choice for $k$ is thus a few orders of magnitude over $p$. The running time of the naive computing procedure is then $O(pn)$ which is worse than Algorithm 8.5. The numerical experiment proposed below confirm this finding.

**Benchmark.** In order to confirm the efficiency of our approach compared to the naïve one, we have been conducting a numerical experiment depicted in Figure 8.5. A sequence of randomized generalized cost PWL functions $g_c$ with an increasing number of pieces have been generated. The generation process is the following. At each step, $k$ Gaussian-shaped PWL functions with random mean are summed. Each Gaussian-shaped function had 50 pieces and its discretization is centred in its mean so the total number of pieces in $g_c$ is $k$ functions times 50. For a given $k$, 20 generalized cost functions $g_c$ have been generated and tested in order to have a stable estimation of the total computing time. The naïve optimization procedure is performed by discretizing the set of preferred arrival times in 200 pieces.

Figure 8.5 shows that the event based approach is always faster than the naive approach. The running time seems to evolve linearly with the number of pieces $n$. According to Proposition 8.2, this would imply that $p$ is either stable or increases linearly with the number of pieces. Yet the noise on the running time curve makes it difficult to confirm the assumption. A closer look at the results reveals that the number of pieces in the arrival time functions is fairly independent from the number of pieces in $g_c$ and varies within 10 to 20 pieces.

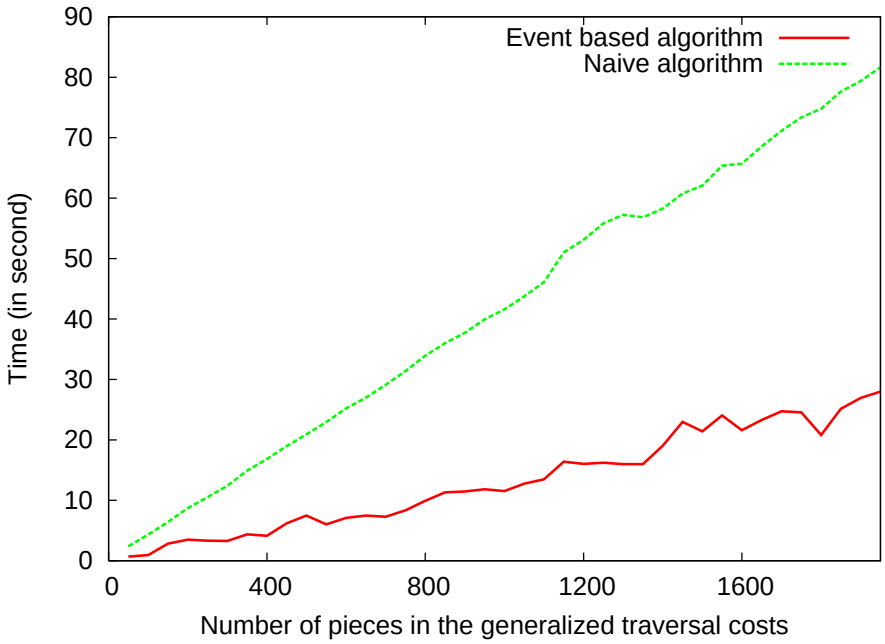



Figure 8.5: Variations of the processing times with the number of pieces in the generalized costs for the naïve and the event based procedures

## 2.3 Extension to convex PWL schedule delay costs functions

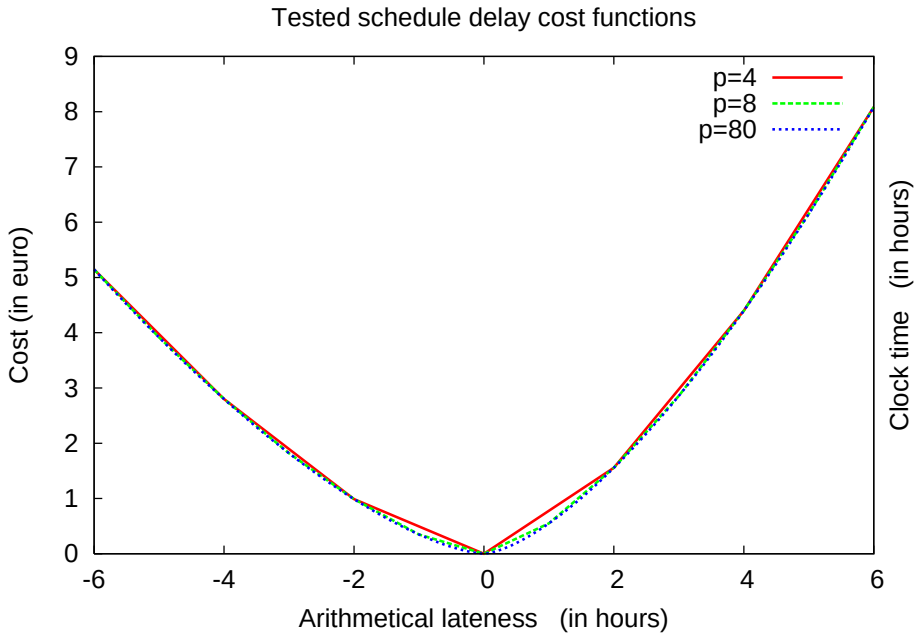
As it has been said in the previous chapter, it is of great interest to consider DUE with more general schedule delay costs functions. There would be several way of extending the previous approach. Most would imply to change the way function are encoded so that the procedure can treat non PWL functions. Yet, for computing reasons, we do not wish to do so. Consequently the approach proposed here is limited to convex PWL cost functions. Consider a continuous and convex PWL schedule delay cost function $D$. The algorithm is quite similar to the one exposed in the previous Subsection. The extension consists in considering several set of candidate instants, one for each pieces of the schedule delay cost function $D$. The rest of the treatment is essentially the same. As its exposition in pseudo-code is rather tedious, it is not done here.

A few examples are given here for successive schedule delay cost functions $D_i$, which are the approximation of the quadratic function:

$$D(l) = \alpha(l^+)^2 + \beta(l^-)^2$$

using PWL functions with respectively 10, 20 and 50 pieces.

The results, shown on Figure 8.6, exhibit an interesting property. Unlike the comparable cases presented in Figure 8.3, the corresponding optimal arrival time functions are "nearly" strictly increasing. The more pieces the schedule delay cost function has, the more this last remark is true. In other terms, there are still constant pieces, but there are much more numerous, hence giving the illusion of a smooth, increasing function.




c

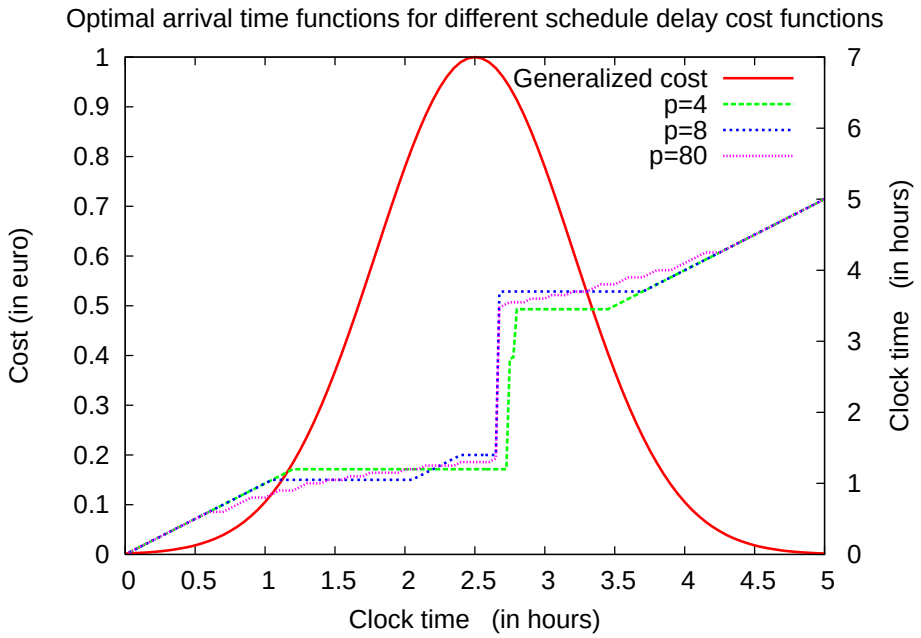



Figure 8.6: A simple illustration of optimal arrival time choice for non V-shaped schedule delay cost functions

This is due to the strict convexity of $D$ or more precisely to the strict monotony of its derivative. Indeed following the same line of reasoning as in Subsection 2.1, one would find that in this case the instants satisfying the first order condition are varying continuously with $h_p$ as soon as both (1) the derivative of $D$ and (2) the derivative of $g$ are varying continuously. None of those two conditions are true, as we have chosen to encode functions using the PWL format. However, when $g$ and $D$ are good approximations of functions actually satisfying (1) and (2), the resulting property also remains "approximately" true.

# 3 Computation of the OD flows from the optimal arrival time functions: the spreading procedure

## 3.1 Motivations

Once for each user's category $c$, an optimal arrival time function $\bar{H}_c$ has been obtained, it remains to deduce $\bar{X}_c$, the corresponding flows at arrival. The most straightforward approach would be to use the following relationship:

$$\bar{X}_c = X_c^p \circ H_c^{-1}$$

Yet, this would lead to discontinuous $\bar{X}_c$, as the functions $H_c^{-1}$ are discontinuous (see the numerical example below in Figure 8.7). Note that this is due to the characteristic assumptions we have retained in our model: flows represented by PWL functions and a V-shaped schedule delay cost function.

However discontinuous $\bar{X}_c$ are not desirable for several reasons. First it causes numerical difficulties in the computation of travel times in the bottleneck model. Second such an approach leads to a non-converging algorithm. A simple interpretation is the following: by concentrating departures at given instants, one only consider discontinuous volumes while we are interested in continuous one. Hence the exploration of the solution space is inefficient. In order to overcome this difficulty, spreading procedures are proposed. They

basically consist in finding all the discontinuities in $\bar{X}_c$ and spread the corresponding volume of user "around" the optimal arrival time.

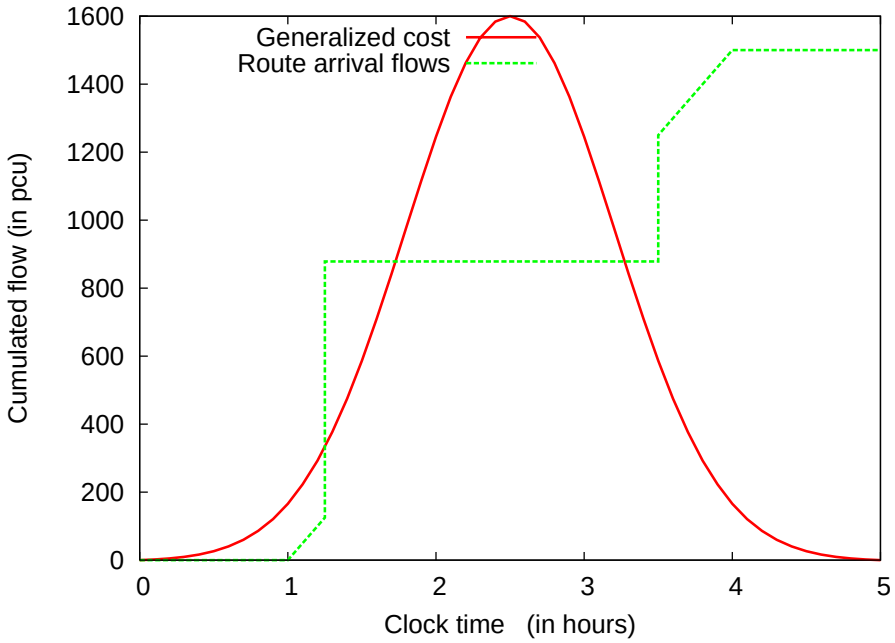



Figure 8.7: Arrival flows before a spreading procedure

## 3.2 Statement

Let us first define what would be an ideal spreading procedure. It is reasonable to expect the following properties:

1. The level of spreading should be affected by the generalized cost function $g_c$.

2. At equilibrium, the spreading procedure should have no effect on the volume.

3. It should be fast to compute.

The two last properties are straightforward, but the first one requires some explanations. Let us take as an example the situation in Figure 8.7. One could simply spread the two volumes corresponding by replacing the discontinuity gap by a piece with a very high slope. In other words, one

could spread the volume using a limit flow thus bounding the maximal flow that can be obtain at equilibrium.

Yet in this approach, one does no to adapt the flows to the current structure of costs. It is natural that when the costs are increasing fast the volumes should not be spread with a low flow, otherwise some users would incur very high costs while other would not. Reversely low variations in the travel cost should lead to high spreading flows.

Thus we propose the following procedure. For each points of discontinuity $h$ in $H_c^{-1}$, compute the interval $I$ such that for all $h_p \in H_c^{-1}(h)$, $h_p \in I$ and optimal route $r'$, one has:

$$|g(h', \tau_{r'c}(h'), p_{r'c}(h'); c, h_p) - \min_{r,h} g(h', \tau_{rc}(h'), p_{rc}(h'); c, h_p)| < dg$$

Then spread uniformly $X_c^p \circ H_c^{-1}(h)$ over $I$. The quantity $dg$ is a positive real parameter and is assumed to be small with respect to the "standard" travel costs. It is referred to as the user's to costs. Intuitively this is stating that beyond a certain difference of costs users are indifferent to two travel alternatives.

Figure 8.8 depicts the application of the spreading procedure on the previous example.

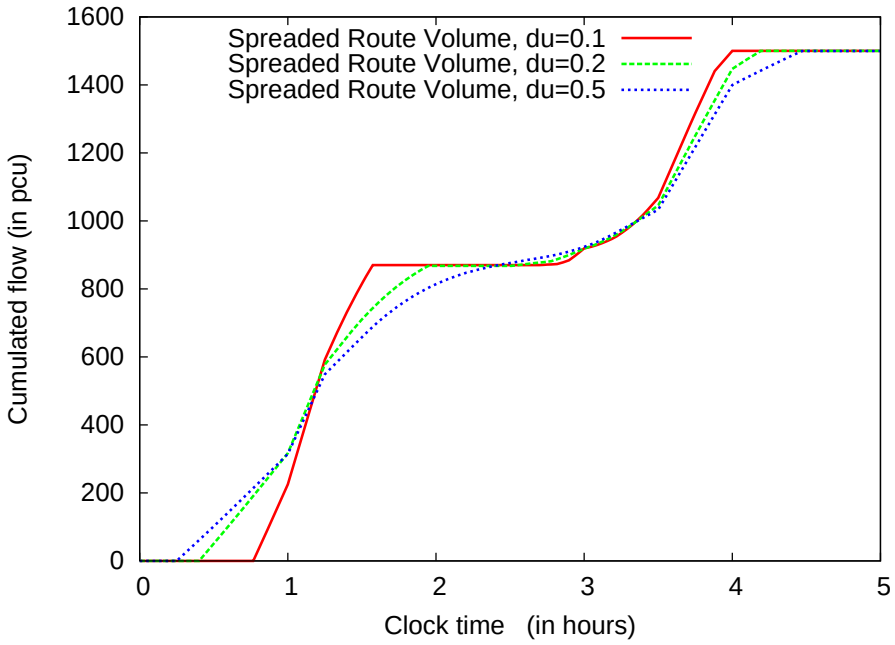



Figure 8.8: Arrival flows after a spreading procedure

# 4 Numerical examples

The algorithm has been implemented in TCL as part of a toolbox called LabDTA (stands for Laboratory for Dynamic Traffic Assignment) presented in Appendix E. The following numerical examples were obtained thanks to this implementation. The prototypes developed in LabDTA were used as a basis to extend the LTK to deal with departure time choice. Some experiments using this latter implementations are presented in Chapter 10.

The first subsection gives the most simple case study: a single OD pair served by a single arc. Therefore the only travel decision of the users regards departure time and no route choice is possible. Moreover we choose uniform users regarding their vehicle type and economic preferences so they are only differentiated by their arrival preferences. The scenario is tested with a V-shaped cost of schedule delay as well as with a strictly convex one.

The second example, called SR91, is a small network with one origin, one destination, two routes and two user categories. Albeit simple, it illustrates combined route and departure time choice. It is also a well known case study

for transportation economists and as such most of the data regarding both scheduling cost functions and value of time are available in the literature.

## 4.1 No route choice

The simple one arc example has already been studied analytically in previous chapters. It is hence especially interesting to see how test the algorithm on those cases in order to compare with the theoretical results.

**V-shaped schedule delay cost function.** The first example we consider is very basic. A single arc with a bottleneck of capacity $k = 20$ uvp/min is subject to a demand where users belongs to $c = (od, (D, \nu), \texttt{vl})$. The distribution of preferred arrival time is $X_c^p(h) = 20.h$ for $h \in \mathcal{H}_p = [0, 10]$ and $D$ is a V-shaped schedule delay cost function of parameters $\alpha = 1.5$ and $\gamma = 0.5$, and the value of time is normalized to $\nu = 1$.

Figure 8.9 shows the cumulated volumes after 50 iterations of the algorithm together with the theoretical solution of the problem. Clearly the computed results are very similar to the theoretical one as far as the cumulated flows are concerned. When the instantaneous flows rather than the volumes are represented (Figure 8.10), the similarity between the two is less obvious. Indeed the convex combination on route volumes used in the algorithm leads to oscillations in the cumulated volumes which in turn are present on the instantaneous flows with an amplification due to differentiation.

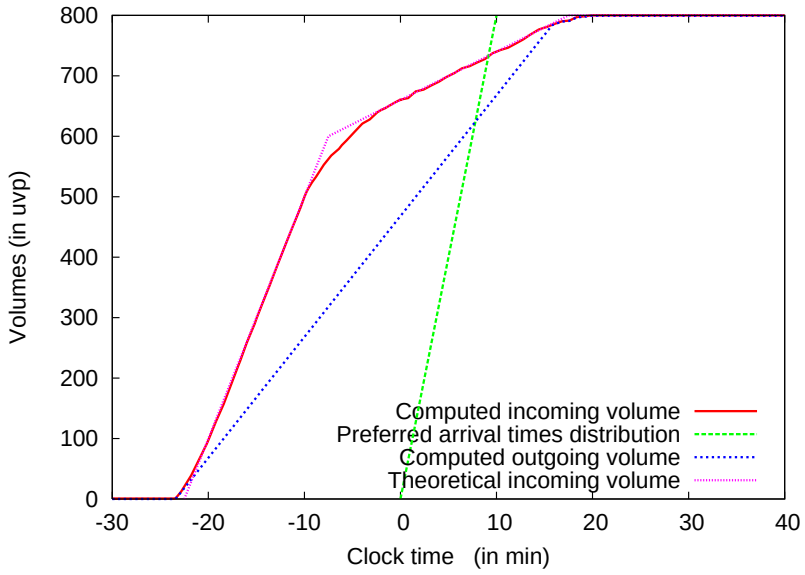



Figure 8.9: Numerical results of the DUE computation algorithm with a V-shaped schedule delay cost function - volumes after 50 iterations.

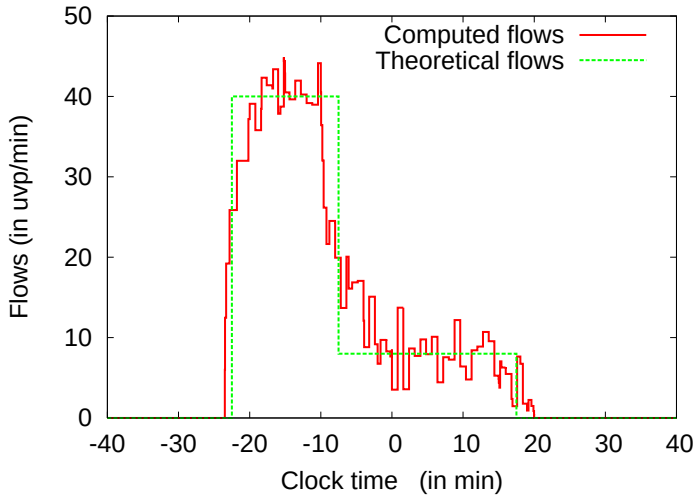



Figure 8.10: Numerical results of the UE computation algorithm with a V-shaped schedule delay cost function - flows

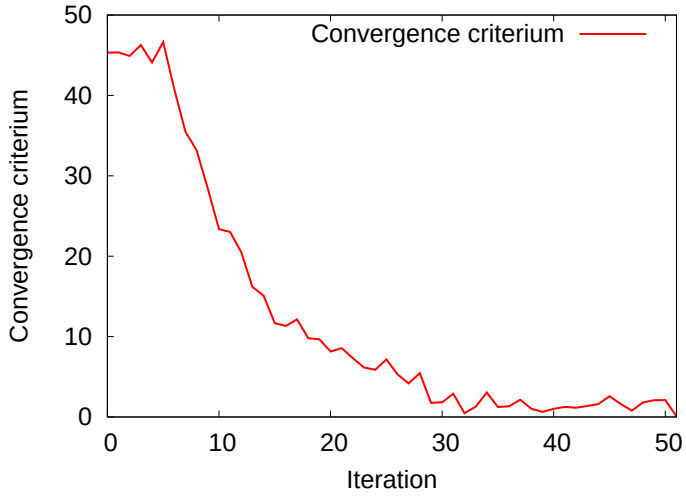



Figure 8.11: Convergence criterion for a V-shaped schedule delay cost function

Figure 8.11 presents the evolution of the convergence criterion with the number of iterations. It depicts a clear converging behaviour with a final value of approximately 0.05. Informally this amounts to say that after 50 iterations, on average users can not lower their costs of more than 5 %.

**Non linear schedule delay cost function.** Now let us apply the UE computation algorithm on the same example except $D$ is now given by a non-linear schedule delay cost function. The chosen schedule delay cost function:

$$D(l) = 40\alpha.\left(\frac{l^+}{40}\right)^{3/2} + 40\gamma.\left(\frac{l^-}{40}\right)^{3/2}$$

The results are presented in Figure 8.12. As in the previous case, the cumulated volumes are qualitatively very similar to the theoretical results, which leads to think the algorithm is converging correctly. This is confirmed by the convergence criterion which is lower than 0.03 after 40 iterations. This is better than for V-shaped schedule delay cost functions and in this case the convergence criterion decreases faster. A possible interpretation is given by the comments made in Subsection 2.3 *i.e.* the less linear a schedule delay cost functions is, the less homogeneous the users' choices are. This eases the computation as the users then naturally spread over the space of departure times.

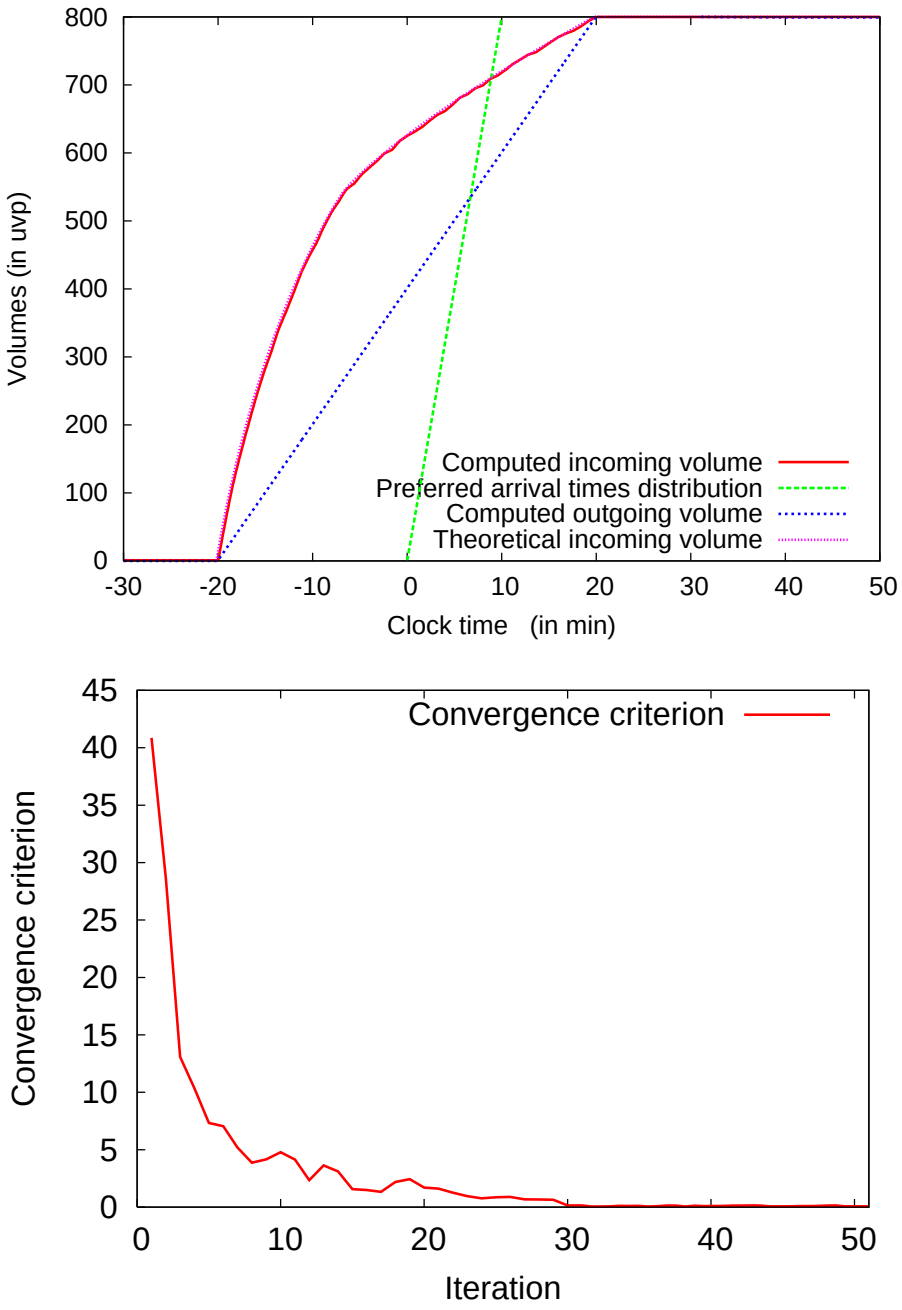



Figure 8.12: Numerical results of the UE computation algorithm with a non V-shaped schedule delay cost function.
- (top) Volumes after 50 iteration
- (bottom) Convergence criterion

## 4.2 The SR91 example

The State Route 91 is located in the orange county (California, USA) and was faced to an important congestion problem at the beginning of the 90's. The road is connecting a residential zone to a labour pool. Before 1995, it had a capacity of 8000 pcu/h, and four lanes in each direction. In 1995 two lanes were added in each direction. The two additional lanes are equipped with time-varying tolls. The tolls are used as a congestion management tool to alleviate traffic on peak hours by spreading demand. Yet this management is complex: a variation in the toll fare can induce both trip rescheduling and rerouting. Commuters are highly heterogeneous regarding their arrival time preference (Sautter, 2007), so the travel time pattern at equilibrium is likely to have a non trivial form.

**User categories.** Two user categories, subscripted $r$ and $p$, are considered. Both of them have a schedule delay cost function $D$ under the classic V-shap form. They differ only by their value of time $\nu$, their unit cost of arriving late $\alpha$ and their unit cost of arriving early $\beta$. For user class $p$, $\nu_p$ is taken from Lam and Small (2001) in its study on SR91. $\alpha$ and $\beta$ are the ratios obtained by Small (1982) in its study for San Francisco. In dollars, the values are $\nu_p = 22.87$, $\alpha_p = 12.20$ and $\gamma_p = 38.12$. For user class $r$, along the line of the technical report from Sautter (2007), it is assumed that the ratios $\alpha_r/\nu_r$ and $\beta_r/\nu_r$ are the same as for $p$, and that $\nu_r = 2\nu_p$.

**Demand and supply.** The distributions of desired arrival instants for each user class are also taken from Sautter (2007) and are estimated on the baisis of historical data from the road operator Cofiroute, in charge of the SR91 since 1995. With these assumptions, there are over 100 000 users evenly distributed between the two categories. The capacities of two routes are set to 8000 pcu/hour for the free route, and 2500 pcu/hour for the two additional lanes.

**Scenarios.** Two scenarios have been simulated. The first is called *untolled*. The amount of the toll fare on the two additional lanes was set to 0, and the user equilibrium with departure time choice has been computed using LabDTA. The second scenario is called *tolled*. The time-varying toll fare was

set equal to the curve plotted in Figure 8.13 (d). A comparative study allows us to discuss the net effect of tolling.

**Results.** The results of each route and each scenario are plotted in Figure 8.13 (a-d). In both cases the results are consistent with the theoretical results from Chapters 5 and 6. In the untolled scenario, the travel time pattern shows a double peak with slopes corresponding to the one induced by the ratios $\alpha_r/\nu_r$ and $\beta_r/\nu_r$. The travel time maxima are obtained when delays are close to 0. Users are using indifferently the two routes and the travel time is rigorously the same. In the tolled scenario, the two users types are segregated by the toll; only users of the category $r$ use the tolled route. Note that the tolling scheme globally reduces travel times. This is achieved by two mechanisms. On the tolled route the toll tends to spread the traffic, thus reducing congestion. On the untolled route, the peak is more spread and slightly less pronounced.

Table 8.1 exemplifies the results of an hypothetical socio-economic analysis based on the results of the simulation. The benefit of the toll is driven by the time savings made by the category $r$, which is natural as they have the entire benefit of the faster tolled lanes. On the contrary users of the category $p$ incur an increase in traversal costs. This increase is nearly compensated by a decrease in schedule delay costs. The explanation is not straightforward. A closer inspection reveals that in the untolled scenario the users of category $p$ are always late w.r.t. to their preferred arrival time. However as there is no toll on their route they tend to schedule their trips regarding to congestion costs, so the benefit of this additional capacity results mainly in a decrease in schedule delay cost.

| | Untolled | | Tolled | | Difference | |
|---|---|---|---|---|---|---|
| Category | p | r | p | r | p | r |
| Traversal Costs | 2947 | 2525 | 3000 | 2432 | -53 | 93 |
| Delay Costs | 1873 | 2923 | 1823 | 2934 | 50 | -12 |
| Total/Category | 4820 | 5448 | 4823 | 5343 | -3 | 81 |
| Total/Scenarii | 10268 | | 10190 | | 78 | |

Table 8.1: Results from an hypothetical socio-economic analysis (values are in thousands of $ per day).

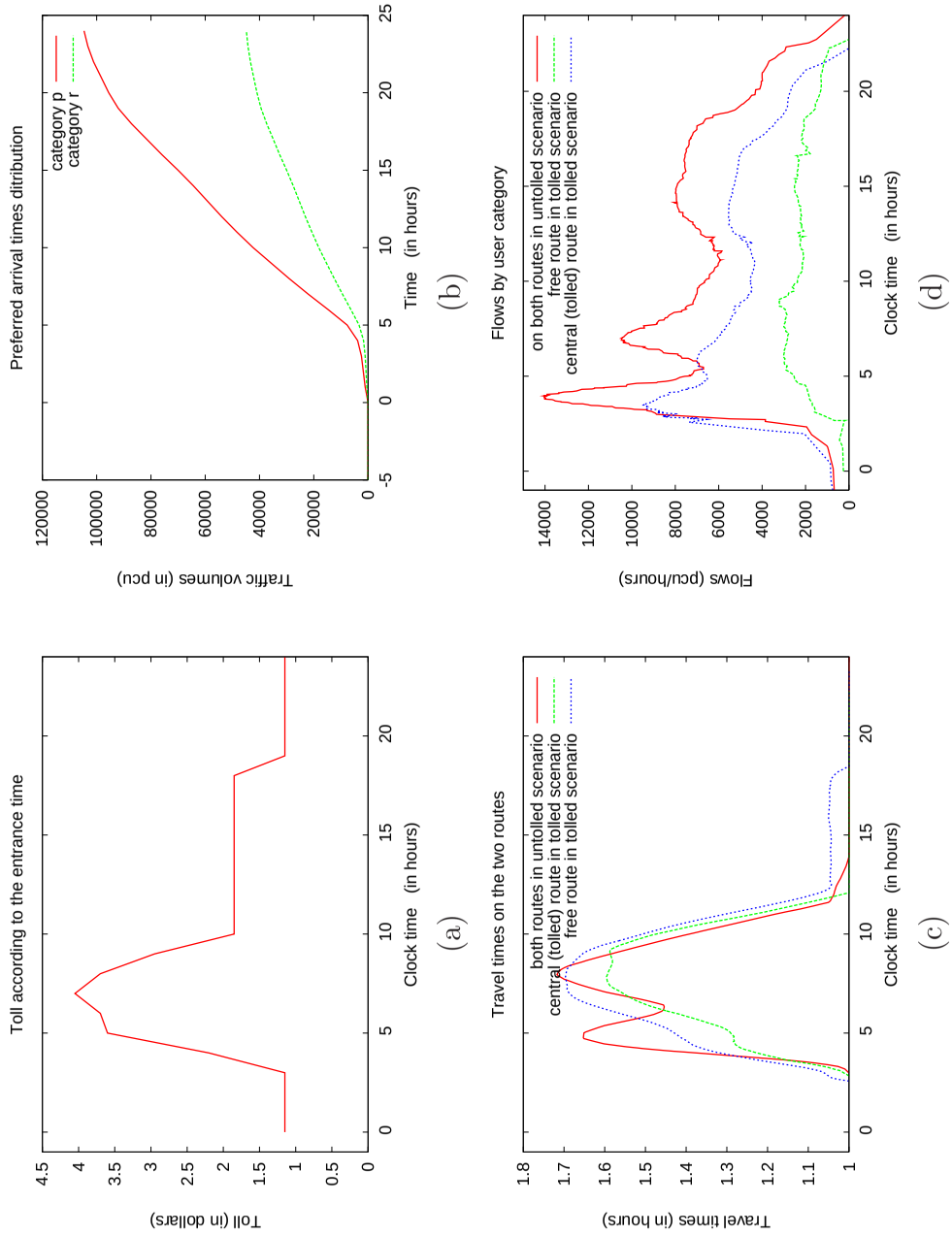



Figure 8.13: Equilibrium values of the SR91 under two scenarios

# 5 Some comments about the user tolerance to costs

The main parameter of the convex combination algorithm is the user tolerance to costs $dg$. In order to achieve good convergence it is necessary to set it correctly. From the numerical experiments that have conducted, we draw the following empirical conclusions:

- The algorithm is *sensitive to the value of* $dg$.
    - *When* $dg$ *is too low* (typically over 0.1% of the average costs) the algorithm tends to stop on a "local minimum".
    - *When* $dg$ *is too high* (typically over 10% of the average costs) the algorithm behaves as a random walk. It seems however that the appropriate $dg$ depends of the level of congestion in the considered case.
- The optimal value of $dg$ varies with the level of congestion.
- *A progressive decrease of* $dg$ *along the execution* of the algorithm improves convergence. Sequences $dg_n = b/(a+n)$ are quite effective. This is the approach retained for the numerical examples.
- PWL approximations of quadratic schedule delay costs functions, as opposed to V-shaped ones, results in situations where the algorithm is less sensitive to $dg$.

Up to now there is no general method to correctly set $dg$. The approach we used in the previous examples was to conduct a sequence of trials guided by the evolution of the convergence criterion over the iterations. Although it leads to reasonable results, this is rather time-consuming.

Chapter 9

# A user equilibrium algorithm based on user coordination

In the previous chapter, an algorithm based on the optimal reaction of users to a state of supply – represented by the travel times and the monetary costs on each route of the network – is presented. This algorithm has proven to be efficient but has some preoccupying flaws. Its efficiency highly depends on the appropriate choice of a parameter, the user tolerance to costs, and there is no systematic approach to correctly estimate it. One is bound to proceed to a sequence of trials which is time-consuming on large networks. The introduction of this parameter is justified by the fact that as users do not coordinate with each other, they tend to choose at the same departure times. Thus the resulting cumulated flows slowly converge toward equilibrium. A spreading procedure parametrized by the previously mentioned *ad-hoc* user tolerance to costs was introduced in order to speed up convergence.

In this chapter, an alternative algorithm that does not rely on a spreading procedure is proposed. It is based on a powerful property that can be informally be stated as such. If all users have taken travel decisions such that they can not decrease their generalized costs by marginally changing their departure time, then the departure distribution describes an equilibrium. This non-intuitive property naturally leads to a more efficient algorithm.

# 1 Context of application and mathematical preliminaries

## 1.1 The restricted model

In this chapter, we consider a specific case of the general model that has been developed in Chapter 7. It is likely that the solution method developed here could be extend to this general model, but it will not be explore here.

The assumptions can be summarized as follows. First, all tolls will be assumed to be null and thus the generalized cost is solely composed of the travel time costs and schedule delay costs. Without loss of generality, the value of time of all users is assumed to be $\nu = 1$, so that the travel time costs can be identified with the travel times. Finally, it is assumed that all users have the same schedule delay cost function and vehicle type. Consequently users only differ by their origin-destination pair and preferred arrival time ; the set of user categories is thus limited to $OD$.

In this context, for any origin-destination pair $od$ the shortest routes to arrive at an instant $\bar{h}$ define a continuous, FIFO travel time function $\bar{\tau}_{OD}$ as a minimum of continuous FIFO route travel time functions. Note that this not the case when they are tolls on the networks. In this latter case the travel times associated to the least cost paths on an OD are not continuous. Frame 3 gives a simple example.

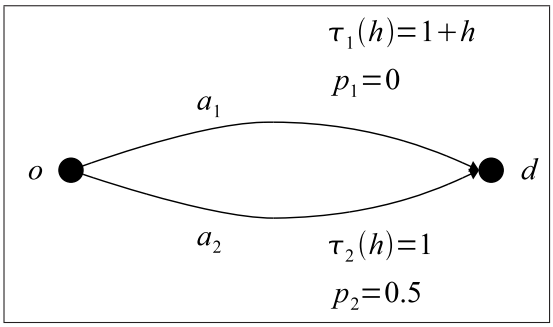
It is then always possible to express it as a function of the departure time rather than of the arrival time, *i.e.* to consider the function $\tau_{OD} := \bar{\tau}_{OD} \circ \left(\mathrm{Id}_{\mathcal{H}} - \bar{\tau}_{OD}\right)^{-1}$. The quantity $\tau_{OD}(h)$ gives the travel time on the optimal route from $o$ to $d$ in order *to arrive at* $h + \tau_{OD}(h)$. Note that $\tau_{OD}$ is well defined as in this case the map $\mathrm{Id}_{\mathcal{H}} - \tau_{OD}$ is inversible. Thus to a given arrival time there is a single corresponding departure time. This is why the problem can be stated from a departure time perspective rather than from an arrival time perspective as done in Chapter 8.

It is interesting to redefine some quantities.

**Supply.** First let us deal with the supply function $F_S$: it now takes as output a sequence of $OD$ flows, denoted $\boldsymbol{X}_{OD}$ and returns a sequence of travel times $\boldsymbol{\tau}_{OD}$ which represents the shortest paths travel times when

Consider the simple two-arc network below and a set of users wishing to go from $o$ to $d$. Assume they have a value of time $\nu = 1$. Then the least cost path is $a_1$ until $h = 0.5$ and $a_2$ for any $h > 0.5$. The travel time on the least cost path written as a function of the departure time is then:

$$\tau_{OD}(h) = \begin{cases} 1 + 0.5h & \text{for } h < 0.5 \\ 1 & \text{for } h > 0.5 \end{cases}$$




Frame 3: Example of the discontinuity of the travel times w.r.t. the departure time on the least cost path

the volumes $\boldsymbol{X}_{OD}$ are assigned at the dynamic Wardrop equilibrium[1]. In other words, $F_S$ is a compact notation for a dynamic assignment algorithm.

**Demand.** Users now only differ from their origin-destination and their preferred arrival time. Thus the set of user categories is limited to $C = OD$ and the set of users is described by a sequence $X^p_{od}$ of distributions on $\mathcal{H}_p$. Their cost function will simply be denoted $g(h, h_p; \tau_{od}) = \tau_{od}(h) + D(h + \tau_{OD}(h) - h_p)$. A sequence $\boldsymbol{X}_{OD}$ such that $X_{od}(\infty) = X^p_{od}(\infty)$ is called an *assignment of the demand*.10

We can now state the dynamic user equilibrium (DUE) in the restricted model:

**Definition 9.1** (Dynamic User Equilibrium in the restricted model)**.** *Find an assignment of the demand* $\boldsymbol{X}_{OD} = (X_{od})_{od \in OD}$ *such that, letting* $H_{od} :=$

[1]Recall from Chapter 2 that the dynamic Wardrop equilibrium is a dynamic user equilibrium with no departure time choice.

$X_{od}^{-1} \circ X_{od}^{p}$ *and* $\boldsymbol{\tau}_{OD} = F_S(\boldsymbol{X}_{OD})$*, for almost every* $h_p \in \mathcal{H}_p$*,* $h' \in \mathcal{H}_+$ *and all* $od \in OD$

$$x_{od}(h) > 0 \Rightarrow g(h, H_{od}^{-1}(h_p); \tau_{od}) \leq g(h', h_p; \tau_{od}) \tag{9.1}$$

Note that in Definition 9.1, the dynamic assignment problem is a subproblem of the DUE computation. This feature is interesting as current dynamic assignment algorithms are already able to treat large scale problems, while DUE with route and departure time choice are still difficult to compute.

## 1.2 The fundamental property

In Chapter 5 an interesting characterization of the DUE has been derived from the first order condition of the user's optimization programs. It has been shown that this characterization was necessary and sufficient and lead to an efficient solution method. In the case of a network (as opposed to the single route case of Chapter 5), a similar proposition can be established.

To do so, let us introduce the concept of departing periods on an origin-destination pair with respect to a given assignment $\boldsymbol{X}_{OD}$. The departing periods $P_1^{od}, \ldots, P_n^{od}$ are the biggest intervals such that $\{x_{od} > 0\} = \cup_i P_i$.

**Proposition 9.2.** *Consider* $\boldsymbol{X}_{OD}$ *an assignment of the demand,* $\boldsymbol{\tau}_{OD} = F_S[\boldsymbol{X}_{OD}]$ *the associated travel times, and denote* $P_1^{od} = [h_1^{od}, h_2^{od}], \ldots, P_n^{od} = [h_{2n+1}^{od}, h_{2n+2}^{od}]$ *the departing period for each od. Assume moreover D is convex and denote* $D_l$ *its derivative. Then the assignment is at equilibrium if and only if:*

$$\textit{for all } od \in OD \textit{ and all } h \in \mathcal{H},$$

$$h \in P_i^{od} \Rightarrow \dot{\tau}_{od}(h) = -D_l\big(h + \tau_{OD}(h) - H_{od}^{-1}(h)\big)\big(1 + \dot{\tau}_{od}(h)\big) \tag{9.2}$$

$$\textit{and for all } od \in OD \textit{ and any } i,$$

$$h = h_i^{od} \Rightarrow g(h, H_{od}^{-1}(h); \tau_{od}) = \min_{h'} g(h', H_{od}^{-1}(h'); \tau_{od}) \tag{9.3}$$

Proposition 9.2 has a simple physical interpretation. Equation 9.2 states that for a user departing at an instant $h$ within a departing period, a marginal variation in the departure time causes no marginal variation in its costs *i.e.* that the function $h' \mapsto \tau_{od}(h') - D(h' + \tau_{od}(h') - H_{od}^{-1}(h))$ admits a local minimum in $h' = h$. The fact that this is necessary condition for equilibrium

is obvious but the other side of the equivalence is quite surprising. Within a departing period, it is sufficient to know that all users are at a local minimum w.r.t. their departure time to guarantee they reached a global maximum on that departing period.

Equation 9.3 states a boundary condition. It concerns users leaving at the boundary of a departing period: their departure time needs to minimize *globally* their cost function.

*Proof of Proposition 9.2.* The "only if" being obvious, let us tackle the "if" part. Take any cumulated flows $\boldsymbol{X}_{OD}$ together with the associated functions $\boldsymbol{\tau}_{OD} = F_S(X_{OD})$ and $H_{od} := X_{od}^{-1} \circ X_{od}^p$ of travel time and preferred time, respectively.

Consider an $od \in OD$. Assume that $X_{od}$ satisfies (9.2) and (9.3). Now consider for any $h$ the function $g^{(h)} : h' \mapsto \tau_{od}(h') - D(h' + \tau_{od}(h') - H_{od}^{-1}(h))$. The quantity $g^{(h)}(h')$ represents the cost incurred by a user of preferred arrival time $H_{od}^{-1}(h')$ when leaving at $h'$.

**Let us show that $g^{(h)}$ admits a global minimum at $h' = h$.** Denote $P^{od} = [h_m^{od}, h_M^{od}]$ the departing period containing $h$. Let us first show that $g^{(h)}(h) = \min_{h'} g^{(h)}(h)$. From its definition $g^{(h)}$ is continuous and differentiable almost everywhere, with derivative $\dot{g}^{(h)}$ given by:

$$\dot{g}^{(h)}(h') = \dot{\tau}_{od}(h') + D_l\big(h' + \tau_{od}(h') - H_{od}^{-1}(h)\big)\big(1 + \dot{\tau}_{od}(h')\big) \tag{9.4}$$

$H_{od}^{-1}(h)$ is an increasing function (as the inverse of an increasing function) and as is $D_l$ because of the convexity of $D$, so the quantity $\dot{g}^{(h)}(h')$ is decreasing with $h$. Around point $h = h'$ we have that:

$$\dot{g}^{(h)}(h') \gtrless \dot{g}^{(h')}(h') \quad \text{if } h' \gtrless h$$

Yet $\dot{g}^{(h')}(h') = 0$ is zero almost everywhere by Equation (9.2), so

$$\dot{g}^{(h)}(h') \gtrless 0 \quad \text{if } h' \gtrless h$$

which means that the function $g^{(h)}$ admits a minimum on $P^{od}$ in $h' = h$.

Then for $h' \notin P$, either $h' < h_m^{od}$ or $h' > h_M^{od}$. Assume there exists $h' < h_m^{od}$ such that $g^{(h)}(h') = \min_{h'} g^{(h)}(h')$. Then as $h_m^{od} = \min_{h'} g^{(h_m^{od})}(h_m^{od})$, Proposition 7.5 of Chapter 8 applies and gives $g^{(h)}(h') = g^{(h)}(h_m^{od}) \leq g^{(h)}(h)$. Using the same arguments for the case where $h' > h_M^{od}$ leads to conclude that the function $g^{(h)}$ admits a minimum on $\mathcal{H}$. Thus $\boldsymbol{X}_{OD}$ satisfies the optimality condition (9.1). □

# 2 The algorithm

## 2.1 General philosophy

An interesting feature of Equation 9.2 is that for a given assignment of the demand it allows to state if the variation of the corresponding travel times is too high or too low at an instant $h$ on an $od$. Informally, the algorithm proceeds as follows. When the travel time variation is too low at an instant $h$ the flow at this instant is increased. If it is too high, it is decreased.

There is a clear relationship between this method and the approach we proposed in Chapter 5 for the one arc case. Equation 9.2 has been derived in the same way as the differential equation of Chapter 5. However, when in the single arc case, the flowing equation gave an explicit expression of $\dot{\tau}_{od}(h)$ as a function of $x_{od}(h)$. There is no such explicit function in the network case. That's why we are going to proceed as if the function $\tau_{od}$ was a black box. When we increase the flow at a time $h$ it is likely to induce an increase in $\dot{\tau}_{od}(h)$, but it is extremely difficult to know of how much; the only way is to test it.

This approach can be related to control engineering. The system is the dynamic assignment procedure, the inputs are the $OD$ flows, the outputs are the travel times and the control law is given by Equation 9.2.

## 2.2 Formal statement

The algorithm is iterative. At every step an update procedure is applied to each $OD$ flows. This procedure modifies them according to the outputs of the system represented by $\boldsymbol{\tau}_{OD} = (\tau_{od})_{od \in OD}$ and the corresponding $\boldsymbol{H}_{OD} = (H_{od})_{od \in OD}$. This is achieved in two parts. First, the demand on each origin destination pair, represented by $X^p_{od}$ is divided in sub-demands $X^{p,1}_{od}, \ldots, X^{p,n}_{od}$ by discretizing the set $\mathcal{H}_p$ in $n$ subintervals. This part of the algorithm is called the `divide` procedure. Then from each subdemands $X^{p,k}_{od}$ are derived new $od$ flows $X'_{od}$. This is called the `coordinate` procedure.

**Division.** The `divide` procedure takes a demand $X^p_{od}$ and divide it according $H_{od}$. This done by choosing a parameter $dh_{max}$ and scanning $H_{od} = (h^p_i, h_i, s_i)$. If one find $i$ such that there is a discontinuity higher than $dh_{max}$ then the demand is divided in $h^p_i$. In PWL format this writes: $i$ is such that $h_i - s_{i-1}.(h^p_i - h^p_{i-1}) > dh_{max}$. The exact algorithm of procedure `divide` is written in pseudo-code below.

**Algorithm 9.1** `divide`$(X^p_{od}, H_{od})$

**Inputs:** A demand $X^p_{od}$, a departure time function $H_{od}$
**Outputs:** A sequence of sub demands $X^{p,k}_{od}$
**Parameters:** $dh_{max}$ the maximum admissible gap

**Set** $h^p_{prev}$ to $\min_{h_p} \mathcal{H}_p$
**For each** $(h^p_i, h_i, s_i)$ in $H_{od}$
  **If** $h_i - s_{i-1}.(h^p_i - h^p_{i-1}) > dh_{max}$ **then**
    **Set** $X^{p,k}_{od}(h) := X_{od}(h)$ for $h \in [h^p_{prev}; h^p_i]$
    **Set** $h^p_{prev}$ to $h^p_i$
  **End For**
  **Set** $X^{p,k}_{od}(h) := X_{od}(h)$ for $h \in [h^p_{prev}; \max_{h_p} \mathcal{H}_p]$

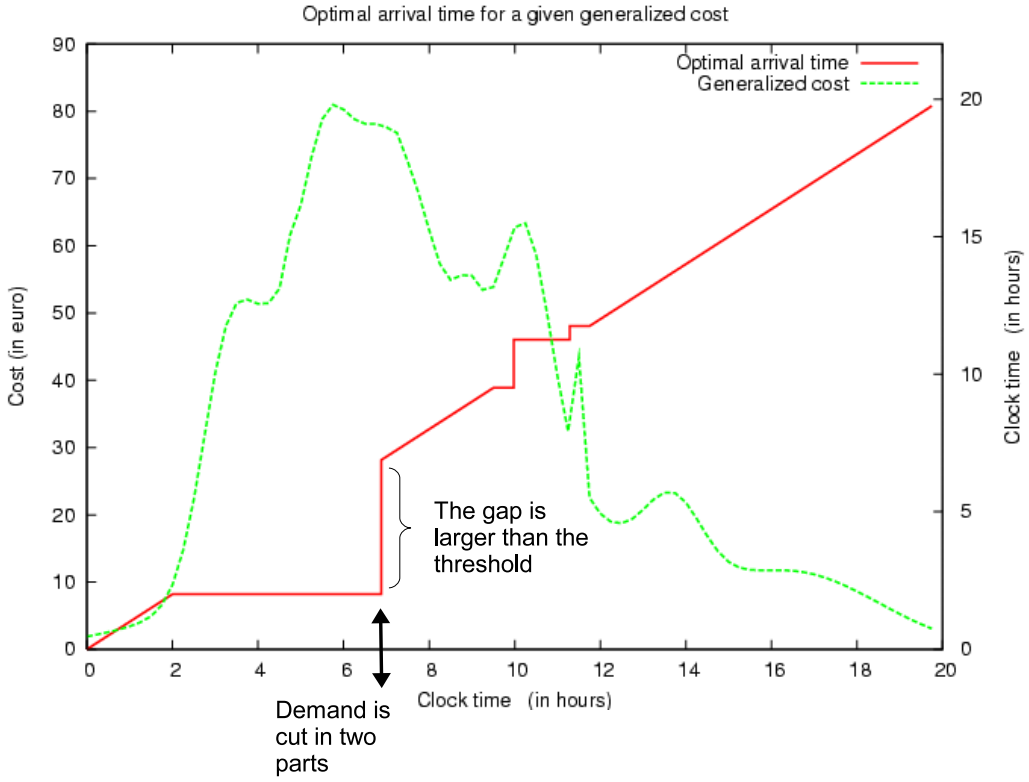



Figure 9.2: Illustration of the `divide` procedure

**Coordination.** The `coordinate` procedure takes a demand $X^p_{od}$ and

assigns it according to the actual travel time $\tau_{od}$ and the previous assignment $X_{od}$.

Define $\eta$ as:

$$\eta(h;\tau_{od},H_{od}) := \Big(\dot{\tau}_{od}(h) + D_l\big(h+\tau_{od}(h)-H_{od}^{-1}(h)\big)\big(1+\dot{\tau}_{od}(h)\big)\Big) \qquad (9.5)$$

The quantity $\eta(h;\tau_{od},H_{od})$ summarizes the gap between the actual variations of the travel time and the one that satisfies the optimality condition stated by Equation (9.2). Note that $x_{od}(h).\eta(h;\tau_{od},H_{od}) = 0$ for all $h$ and $od$ is a necessary condition for the equilibrium.

The first step is to find the optimal departure instant $h^M$ for the preferred arrival time $h_p^M = \sup\big\{h_p : x_{od}^p(h_p) > 0\big\}$. Note that $h_p^M$ can be interpreted as the user with the highest preferred arrival time. Then the new cumulated flow $X'_{od}$ is obtained by the solving the following functional system:

$$\left\{\begin{array}{rcl} X_{od}(h) & = & X'(h^M) - \displaystyle\int_{H'_{od}(h)}^{H'_{od}(h^M)} \max(\alpha\eta x_+,0)dH'_{od} \\ H'_{od}(h) & = & \big(X_p(h^M)-X_p(h)\big)^{-1} \circ \big(X'_{od}(h^M)-X'_{od}(h)\big) \end{array}\right. \qquad (9.6)$$

Algorithm 9.2 explicits a discretized version of the procedure presented above. The right approach here would be an exact resolution of (9.6), so it is merely given here to ease the understanding of the procedure, rather than for a direct implementation.

**Algorithm 9.2** `coordinationProc`$(X_{od}^p, \tau_{od}, X_{od})$

**Inputs:** A demand $X_{od}^p$, a travel time $\tau_{od}$ and a cumulated volumes $X_{od}$
**Outputs:** A new volumes $X'_+$
**Parameters:** $x_m$ the minimal admissible flow
Discretize $X_p$ into a sequence $((h_p^1,V_1),\ldots,(h_p^n,V_n))$
Consider $h_p^n$ and **Set** $h_n$ to the associated optimal departure time w.r.t. $\tau_{od}$
**Set** $X'_{od}(h_n) := \sum_k V_k$
**For** $i = n-1,\ldots,1$ **do**
  **Set** $x_+^k := \max(x_+(h_{k+1}).\big(1+\alpha\eta(h_{k+1})\big), x_m)$ where $\eta$ is given by (9.5).
  **Set** $h_k := h_{k+1} - V_k/x_+^k$ and $X'_+(h_{k+1}) := V_k$
**End For**

# 3 Numerical example

The user coordination algorithm has been implemented in LabDTA (see appendix). The result on the simple one-arc network of Chapter 8, Section 4 is presented and assessed in this section. The numerical setting is rigorously the same. The stress is on the comparison with the convex combination algorithm.

Figure 9.3 presents the cumulated flow representation of the results after 20 iterations. Graphically the results perfectly fit the theoretical ones. The convergence criterion is under 0.01 whereas with the convex combination algorithm it was 0.1. Moreover Figure 9.3 shows that the convergence is much faster in terms of iteration. Let us add that each iteration is faster to compute, so at the end the algorithm with user coordination is especially efficient on this simple case.

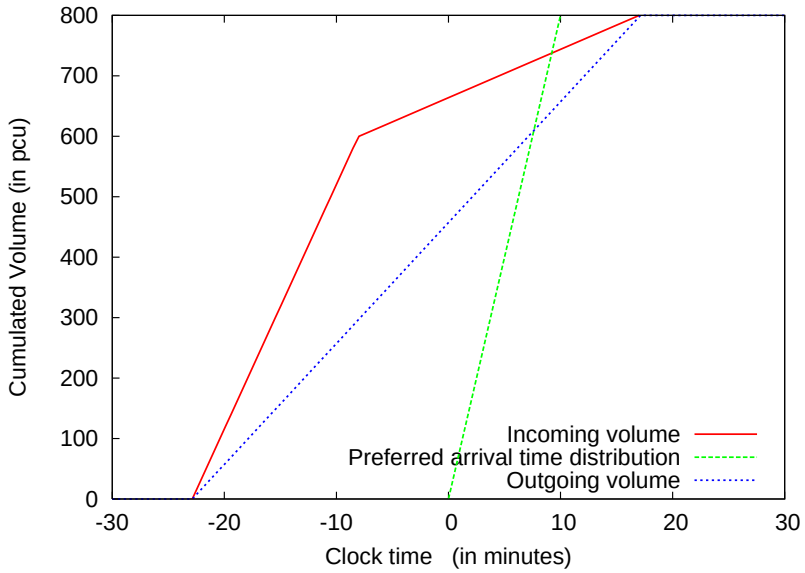



Figure 9.3: Results of the user coordination algorithm on one arc after 20 iterations

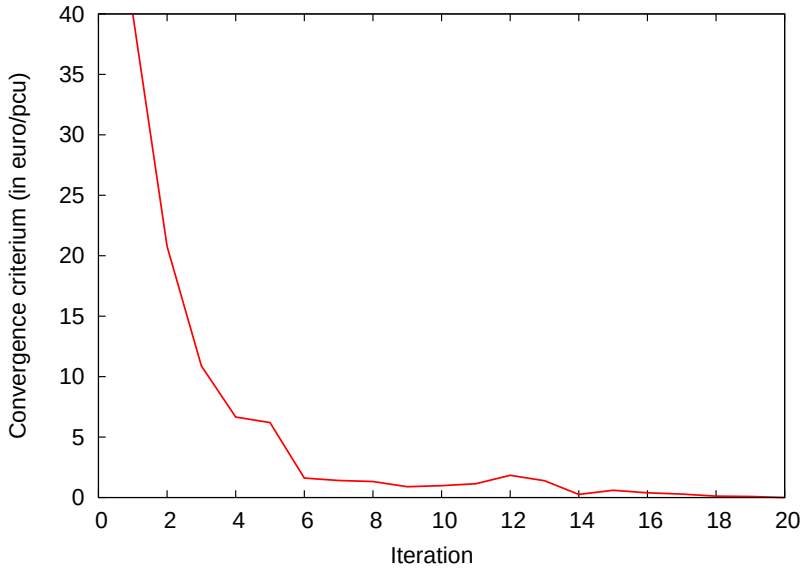



Figure 9.4: Convergence of the user coordination algorithm

# Conclusion

This chapter has presented a proposal for a new algorithm for the Dynamic User Equilibrium computation. The algorithm was tested and assessed on a small example and the results are especially encouraging.

The algorithm remains to be tested more extensively to control how it behaves on larger networks. Moreover it would be interesting to see how to extend it on wider case of applications and especially to see how to take in account tolls on the network. This gives perspectives for future works.

Chapter 10

# An application to a large interurban network during summer holiday departures

During summer holidays, a significant part of the trans-european road traffic is concentrated in the Vallée du Rhône (VDR) area. Tourists coming from northern Europe (including Belgium, the Netherlands, Germany and Great Britain), travel across France to reach (or return from) southern countries (*e.g.* Italy and Spain), meeting on their way people from the Paris area. The situation is depicted in Figure 10.1. The map shows the location of the VDR area and sketches the structure of traffic flows from foreign countries. The main axis in the VDR area in the A7 motorway, located between Lyon and Orange. The distance between those two cities is around 200km. During summer Saturdays, traffic conditions on motorways are usually very bad, especially on the A7, because of high levels of congestion.

To better operate the network, motorway operators have shown interests in studying time varying tolling strategies. Among the possible schemes, a toll varying *within the day* and *from day to day* is especially appealing for summer holidays trips as it enables the operator to influence the departure day as well as the departure time. The aim of this chapter is to assess such a strategy.

Results presented in the sequel illustrate the ability of our model and the associated algorithms to handle such kind of studies on large networks and to give reasonable orders of magnitude. However it does not intend to show the algorithm convergence on large networks or to provide accurate prevision of the traffic level. This issue is state-of-the-art research problems and an

important amount of work is still required before designing effective methods to compute DUE with departure time choice on large networks.

This chapter is divided in three sections. First, it is argued that the model of Chapter 7 is suited to inter-urban travel. Second, the details of the numerical set up are presented. Finally, some numerical results are presented and commented.

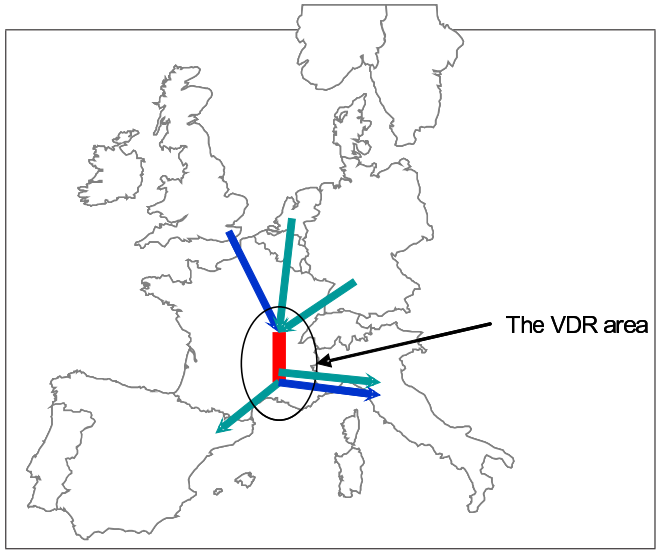



Figure 10.1: Vallée du Rhône (VDR) location and main traffic flows during holiday departures.

# 1 Empirical evidences on inter-urban travel and their practical implications for departure time choice modelling

Dynamic user equilibrium model with departure time choice essentially draw upon Vickrey's bottleneck model (see Chapter 1). Two assumptions underlie the bottleneck model and its various extensions for networks: (1) preferred arrival times are taken from a discrete set of values and (2) delay cost functions are convex. As discussed in this subsection, those two assumptions appear not to be appropriate means of modelling economic preferences of inter-urban trip makers.

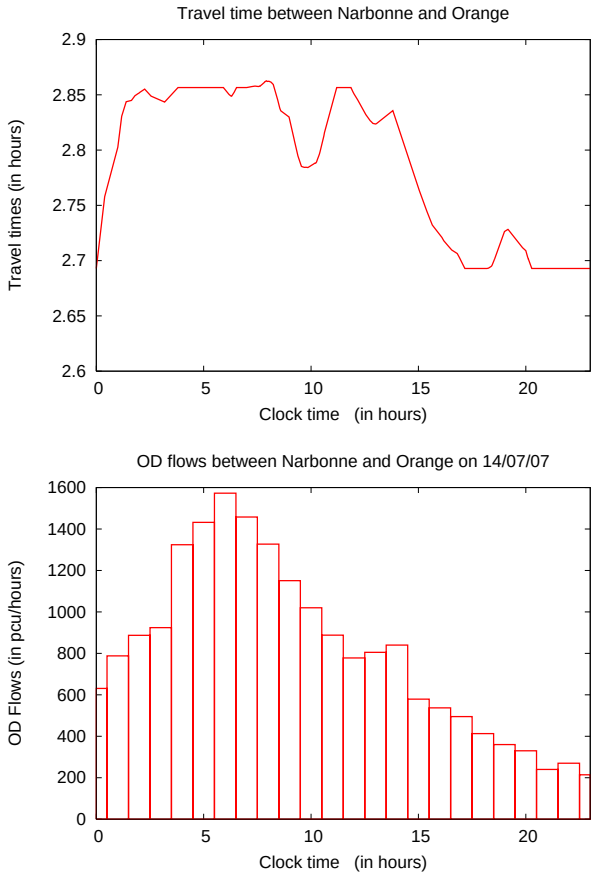



Figure 10.2: (Top) Travel time pattern between Narbonne and Orange on the 14th of July 2007. (Bottom) OD flows between Narbonne and Orange (courtesy of ASF).

Let us first have a look at the travel time patterns and flow rates observed in inter-urban trips. Figure 10.2 (Top) shows the variations of travel time between Narbonne and Orange, two cities of southern France, on a holiday departure day. The pattern is quite far from the single peak predicted by the bottleneck model: at least two peak periods can be observed, along with significant variations elsewhere. The flow rate on the same OD pair is plotted in Figure 10.2 (Bottom). One can observe significant flow rates during the whole day. This is clearly not consistent with a single preferred arrival instant.

Another interesting point is the diversity of inter-urban travellers. As opposed to the morning peak where the road traffic is mainly composed of commuters, inter-urban trips have a wide variety of purposes, inducing significant differences in value of time and delay cost functions. In the same order of idea, a significant part of the traffic is composed of heavy vehicles, which has important consequences on congestion modelling.

Finally the results of a survey organized in 2008 by three French motorway operators show that, during summer holidays, trip makers can be divided into two categories:

- **Unflexible users** can absolutely not afford arriving later than scheduled (*e.g.* because they need catching the key for their rental) and are not ready to change their day of arrival.

- **Flexible users** are far less constrained at arrival. They may change their day of arrival or/and arrive after their preferred arrival time. They are even ready to reschedule their departure day, if they can benefit from lower congestion or toll fares.

This last point is especially interesting as it shows that in inter-urban context the convexity of delay cost functions can no longer be assumed. Indeed in this case the cost of the delay does not necessarily decrease as the arrival time gets closer to the preferred schedule. A traveller considering to leave one day in advance to avoid traffic jams will not necessarily consider arriving at 2 a.m. a better option.

To sum it up, empirical observations show that, for inter-urban trips, a departure time choice model should differ from the classical "bottleneck-like" approach, with respect to the three following requirements:

1. A high level of heterogeneity regarding both preferred schedules (several

preferred arrival times per OD pair) and economic characteristics (value of times and schedule delay cost functions) is required.

2. Multi class congestion modelling is to be considered.

3. Users should be able to choice their day of departure as well as the time of the day.

This justifies the use of the modelling framework presented in Chapter 7 in this case. The following section presents the model set up.

# 2 Model set up

## 2.1 Modelling details

Two user classes, named *const* and *flex*, were considered. Those two classes correspond to passenger cars[1]. They share most of their characteristics (same free flow travel time, same toll prices,...). They are distinguished only by their schedule delay cost functions. For user category *const*, the penalty of arriving later than scheduled grows linearly at a very high rate, while the penalty of arriving sooner grows at a lower rate. The schedule delay cost function of user category *flex* is a little bit more complex. Around 0, it has a classical V-shape, except that the cost of a early arrival grows faster than the cost of a late arrival. Between 6 and 18 the cost of the delay is infinite. Around 24, the shape of the delay cost function is similar than around 0, except that it is shifted up by an amount that corresponds to the cost of rescheduling the departure to the day after. The values of the delay cost evolve similarly around 24. This expresses the cost of rescheduling the departure to the day before. The schedule delay cost function of user category *flex* is depicted in Figure 10.3.

[1] The traffic of heavy vehicles is ignored since traffic regulation rules forbid truck traffic during some of the most congested days in summer.

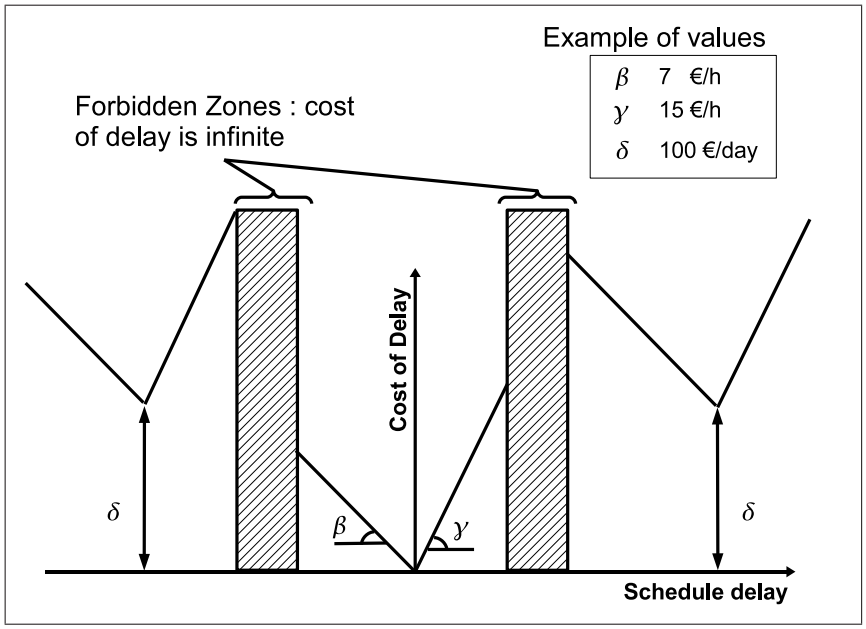



Figure 10.3: Schedule delay cost of the user category *flex*. The values given here are just orders of magnitude. The actual values used for the simulation are not given for confidentiality reasons.

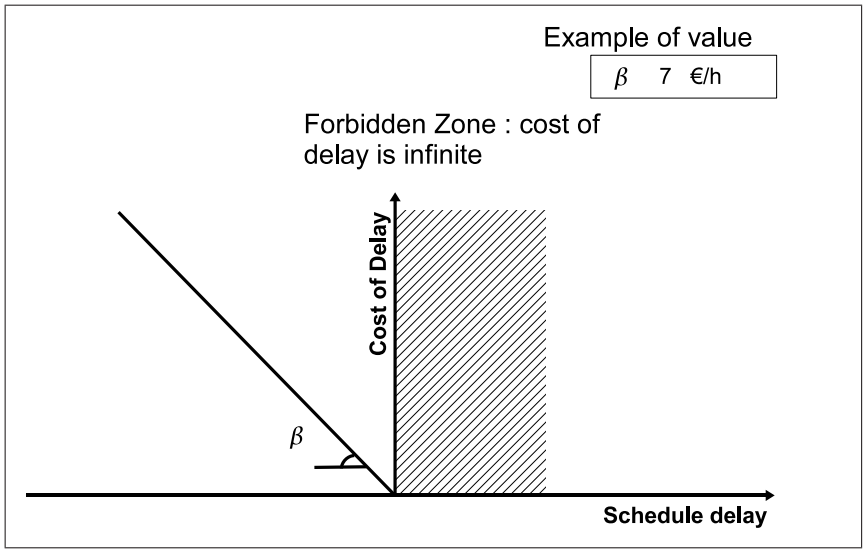



Figure 10.4: Schedule delay cost of the user category *const*. The value of $\beta$ given here are just orders of magnitude. The value of $\gamma$ is taken sufficiently high, that it is numerically close to infinity. The actual values used for the simulation are not given for confidentiality reasons.

## 2.2 Input data and calibration

Most of the data was provided to us by courtesy of companies of the Vinci Group (ASF, APRR and Cofiroute) operating the motorway network in the area of interest. The network comprises 2404 arcs and 939 nodes and is depicted on Figure 10.5. We have for each arc its capacity, free flow travel time for passenger cars, and travel price for passenger cars. The demand is expressed for 628 OD pairs. The simulated days are July the 14th and the 15th, 2007.

The tested scenario is the introduction of time-varying tolls on the motorways A5, A54, A6, A7, A9, A10, A11 and A71. They are plotted in red in Figure 10.5. The time-varying tolls have been built by multiplying the former tolls by a time dependent factor, which is plotted in Figure 10.6. This factor is greater than one (i.e. the fare is higher than usual) between 5 and 17. Its is lower than 1 between 20 and 34, meaning that the amount of the fare is lower than usual between 8 p.m. of the simulated day and 10 a.m of the day after. Although the network is pretty important, we will focus solely on the VDR and its surroundings.

We also have at our disposal a distribution $\boldsymbol{X}_c$ of the preferred arrival times for all OD pairs and users categories. It has been inferred for the simulated day from a large survey conducted by the motorway operators in year 2008. The value taken for the parameters are also inferred from this survey.

Most of the motorways of the network under study are equipped with closed toll systems. As a consequence, we had at our disposal an accurate time-dependent OD matrix for the simulated day, built from the toll stations records. This allows for a fine grain calibration of the model, by adjusting its parameters until simulated traffic flows computed by a (fixed demand) traffic assignment match well traffic counts data, for a significant percentage of motorway sections. A calibration was performed by the Economy and Traffic Department at Cofiroute with very encouraging results.

However the results presented here are using an uncalibrated set of data for confidentiality reasons. Thus the sole purpose of the results presented below is illustrative, and the conclusions, figures and charts presented therein have *no particular meaning outside the scope of this thesis.*

## 2.3 Computation

The computation algorithm used here is essentially the method presented in Chapter 8. A slight change has been introduced to deal with problems of large size. At each iteration the arc cumulated flows are approximated by reducing the number of pieces that describe them. The approximation is due to Aguiléra (2010). More details may be found in this latter reference. The convex algorithm has been implemented in the Ladta Toolkit from the prototype developed in Chapter 8. A computer equipped with a bi-core processor and 2Gb of RAM were used. A total of 50 iterations were performed, yielding a total run time of roughly 1 hour.


Figure 10.5: Network under study. The motorways where the time-varying tolling scheme is implemented are plotted in red.

# 3 Results

## 3.1 Comments about convergence and computation times

Figure 10.7 (Top) depicts the inter-iteration criterion over the iterations of the algorithm. The inter-iteration criterion is defined as such

$$I_k = \sum_{a \in A} \frac{||Y_a^{[k]} - Y_a^{[k-1]}||}{Y_a^{[k]}}$$

where $Y_a^{[k]}$ is the cumulated flow on arc $a$ at iteration $k$ and $||.||$ is the supremum norm.

This kind of criterion has well known flaws. In particular for a method of successive averages, it converges by construction towards 0. In this case it was difficult to use anything else. A finer criterion, like the one developed in Chapter 7, would require the route flows or at least a decomposition of the arc flows by destination. The memory requirements for this are too important on a network of this size. Now, even if the inter-iteration criterion is not suited to state precisely on the quality of the equilibrium, it is an adequate stop criterion. When $I_k$ becomes close to 0, it is not interesting to carry on the algorithm as the current solution is not going to evolve much by performing some extra iterations. Figure 10.7 (Top) shows that we have performed enough iterations.

Figure 10.7 (Bottom) depicts the computation time. An initial increase happens after the first iteration and a more gentle decrease is observed for the rest of the computation. As already mentioned the overall computation time is very reasonable (approx. 1 hour).

To sum up it is difficult to measure the exact quality of our solution here, but the elements exposed above allow us to say that our algorithm can be applied to reasonably large networks without any operational difficulties.

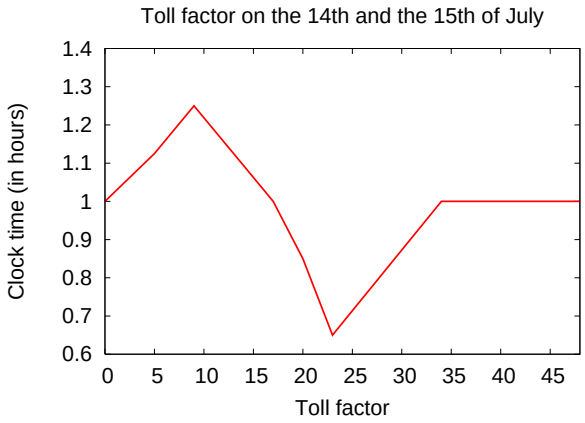



Figure 10.6: Toll factor

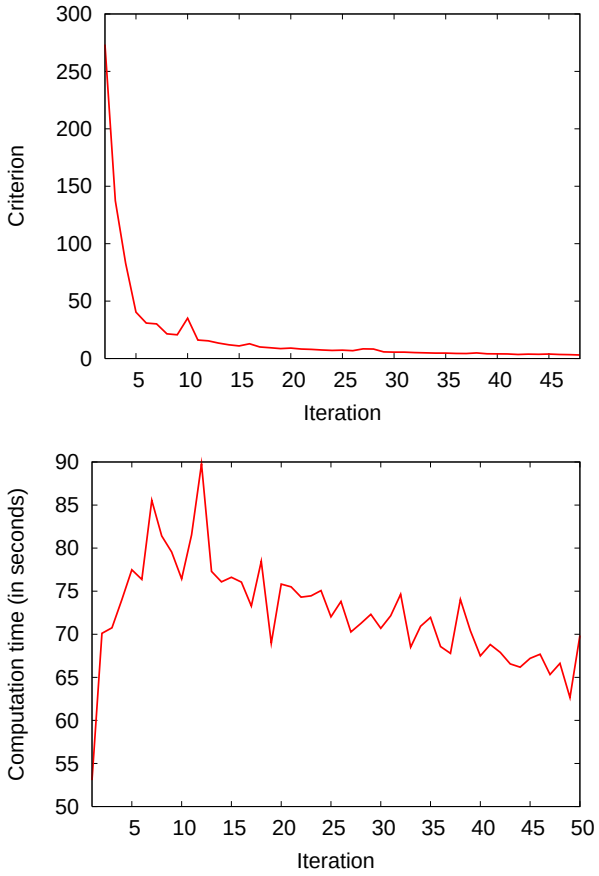



Figure 10.7: Convergence results

## 3.2 Aggregated results

The results are presented for two scenarios. The scenario with *time-invariant* tolls is called the current situation, while the scenario with *time-varying* tolls is called the projected situation.

Figure 10.8 shows the overall distribution of departure times for both scenarios. It illustrates the shift of departure time caused by the time-varying tolls. This occurs at two levels. Within each days, users' departure times are distributed more evenly. Between days, more users depart on the second day in the projected situation than in the current one.

Figure 10.9 (Top) represents the following congestion indicator:

$$\omega(h) = \sum_{a \in A} x_a(h).\tau_a(h)$$

Note that $\int_{\mathcal{H}} \omega(h)dh$ is the travel time aggregated among users, so $\omega(h)$ can be interpreted as a variation of this quantity over time. $\omega(h)$ is a useful indication of the temporal repartition of the aggregated travel time. Figure 10.9 (Top) confirms users' shift of departure time as the travel time are more evenly distributed over the two days. The integration of $\omega(h)$ for the two scenarios also show the aggregated travel time decreases from approximatively 10% in the projected situation.

Figure 10.9 (Bottom) represents the following congestion indicator:

$$\omega^q(h) = \sum_{a \in A} x_a(h).(\tau_a(h) - t_{a,0}(h))$$

It is the equivalent of $\omega$ for the the time spend queuing. Figure 10.9 (Bottom) shows that in the projected situation some congestion appear on the second day.

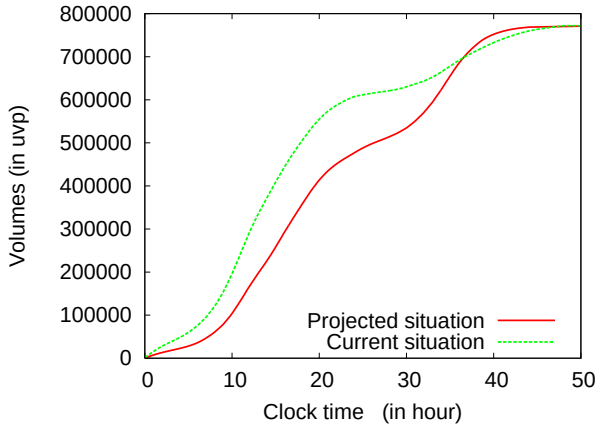



Figure 10.8: Departure time cumulated distributions

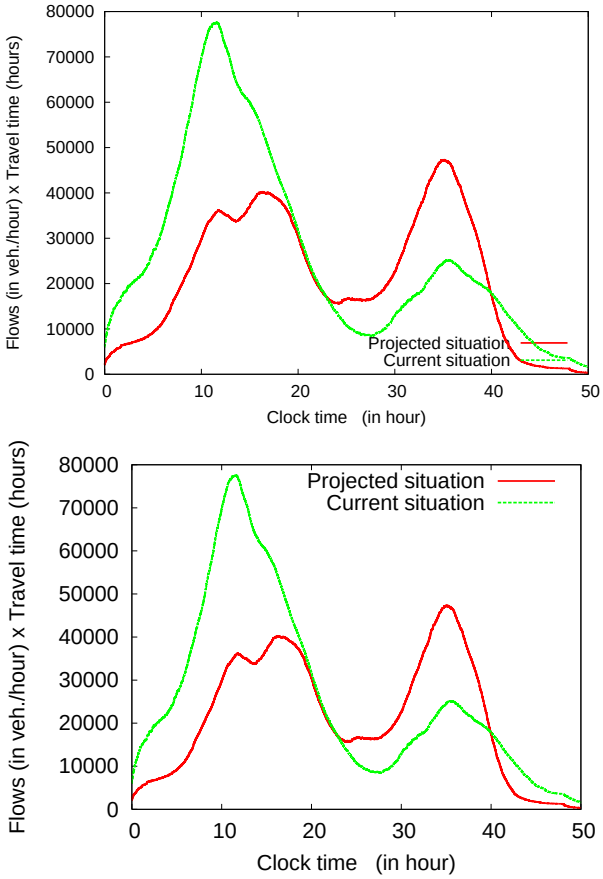



Figure 10.9: Travel times (top) and queued times (bottom) on the French road network

## 3.3 Disaggregated results

Figure 10.10 shows a map of the arc travel times at 11:00 am on the 14th of July in the current scenario while Figure 10.11 shows the same map for the projected scenario. Both maps focus on the VDR. The projected scenario is clearly less congested. An interesting point is that some congestion appears on secondary roads, in particular on the N88 road going from Saint Etienne to Le Puy-en-Velay. This is caused by the users' shift from tolled roads to untolled ones. It thus reasonable to state that the congestion decreases is caused by two mechanisms: the spread of the demand in time, both within and between days, and a spread of the demand among the routes. In the projected situation the road capacity is used more efficiently both in time and space.

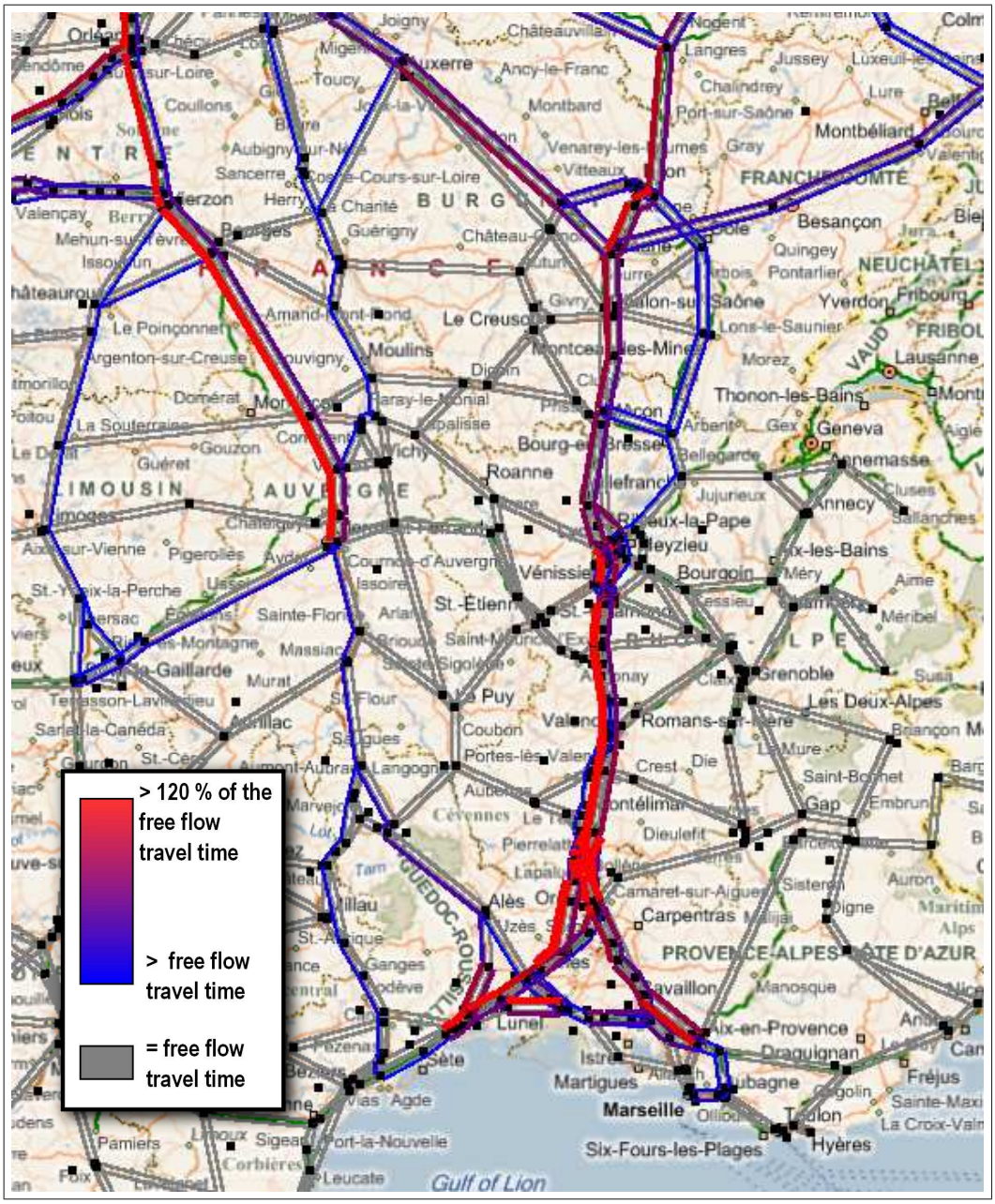



Figure 10.10: Travel times in VDR in the current scenario

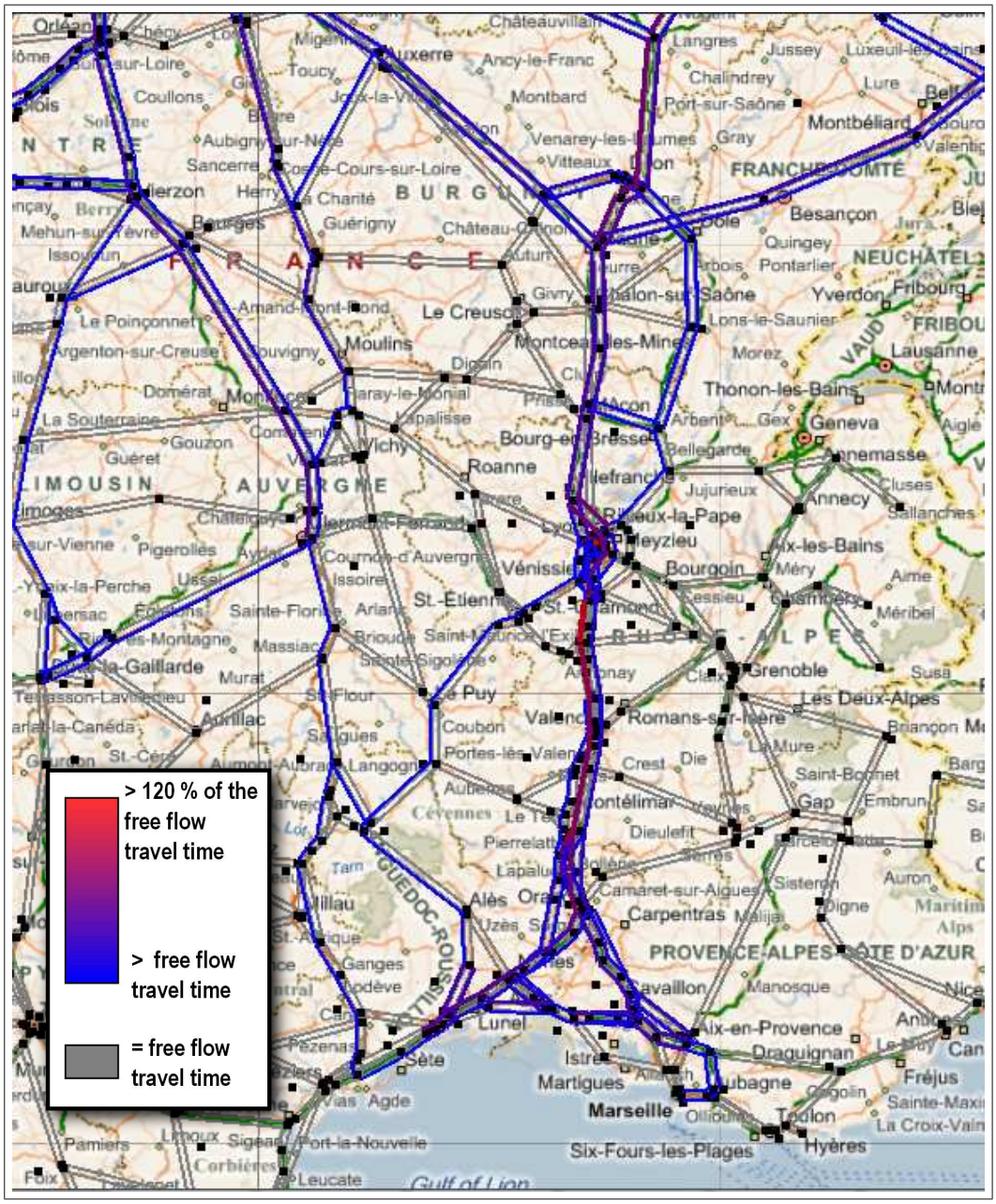



Figure 10.11: Travel times in VDR for the projected situation

## Conclusion

This chapter exposed a possible application for the model developed in this part of the thesis. The relevance of the assumptions has been demonstrated and the computational burden associated with a network of this size is perfectly acceptable. The qualitative analysis of the results showed the computed equilibrium gives consistent orders of magnitude. Now a proper assessment of the convergence of the algorithm on large scale networks would require to compute a rigorous criterion. For computational reasons, it was not done here. This leaves interesting perspectives for future works.

The numerical results show that a moderate toll-varying scheme might have strong impacts on an highly congested interurban network. By applying a time-dependant factor varying between 0.7 and 1.2 to the existing tolls, the aggregate travel times have decreased of approx. 10%. The resulting congestion decreases is apparently caused by two mechanisms: the spread of the demand in time, both within and between days, and a spread of the demand among the routes. It is important to note that this was achieved using a limited increase relatively to the users' generalized costs.

# General conclusion

# General conclusion

This work is purported to provide a game theoretic analysis of dynamic user equilibrium models. The approach can be broken in three steps. First a general framework, the dynamic congestion games, has been set up. Then, two specific dynamic congestion games for which the solution can be derived analytically are studied. Finally, we state a dynamic user equilibrium model in the formalism of dynamic congestion games and present numerical approaches to solve it.

This set of selected issues have been addressed, yielding specific answers to specific questions. Brought together these results allows to give elements of answers to the questions raised in the introduction. These conclusions are outlined here.

**Towards a unified framework for dynamic user equilibrium models.** Since the seminal work of Friesz et al. (1993), DUE models of an increasing complexity have been proposed, and yet the transport science community is still in search for a unifying framework. Chapter 5 and 6 have highlighted the importance of continuous user heterogeneity in the representation of transport demand and exposed how to model it on small networks. These models can be seen as special cases of a wider framework, dynamic congestion games, that has been presented in Chapter 3. The relationship between dynamic congestion games and dynamic traffic assignment models is explored (Chapter 4) and it is shown that the standard deterministic route choice approach to dynamic traffic assignment can be formulated as a dynamic congestion game.

Now, as discussed in the conclusion of Chapter 3, dynamic congestion games may encompass a much wider range of DUE models *e.g.* models with departure time choice, with distributions of the value of time and possibly

simple activity-based models. In a nutshell, they seem to provide an interesting framework for equilibriums with complex demand representations. Carrying on the investigation of the relationships between dynamic congestion games and existing DUE models in that direction is, to the author's viewpoint, an interesting continuation of the research conducted in this thesis. A particularly interesting outcome could be existence results for DUE problems with departure time choice, for instance the one of Friesz et al..

**About the continuous time approach.** In the previously mentioned paper of Friesz et al., the authors stated that "*time is experienced as a continuum and should be modelled that way*" (p 189). We share this opinion, and this was one of the fundamental modelling choice of the LADTA model. This thesis explores at least two consequences of this choice regarding DUE models with departure time choice.

The first issue is the *representation of user choices*. In Chapter 5, we derived a model where preferred arrival times and departure times have a continuous representation. This leads us to investigate the complex issue of correctly representing user choices over a continuous set of decision variable (the departure times) when users are themselves distributed over a continuous set (the preferred arrival times). This specific approach is extended to a much more general case in Chapter 3. This rigorous representation of the user choices proved to be insightful in Chapter 7 and 8, where it leads to design a reduced formulation of a DUE problem with departure time choice, that is both well-suited to computation and offers some existence guarantees.

The second issue is the computation of *deterministic DUE with continuous times*. As noted in our literature review (Chapter 2), most of the current DUE models assume stochastic departure time choice and it seems that this behavioural assumption is motivated by computational reasons. Deterministic DUE computation is the main object of Part IV. Computation algorithms are presented and rigorously assessed on small examples. The operational performance of one of these algorithms is then tested on a real life network.

**The difficulties related to time versus costs DUE.** Replacing travel times by generalized costs in the formulation of a DUE might sound straightforward, but in fact leads to conceptual and algorithmic difficulties. In our literature review (Chapter 1), we exposed that the shortest path problem exhibits very interesting properties, especially from a computational per-

spective, while most of them were not true for least cost path problems. For instance, a least cost path on a network might not be loop free. This has consequences on the way one formulate and solve DUE models. In Chapter 7 and 8, this leads us to consider the user decision strategy as a two-stage process: first, the arrival time choice and then the route choice. The last numerical experiment in Chapter 8 also highlights the increased computational difficulties and instabilities observed when considering generalized costs rather than travel time-based equilibriums.

**Open issues.** It is commonplace for research works to give some answers, but more questions ; the present one is no exception to this rule. Some possible future works are listed below:

- *The existence of an equilibrium in Friesz et al. (1993)'s model* has not been addressed yet. One might wonder whether dynamic congestion games are a suitable concept to achieve this result.

- *The extension of dynamic congestion games to encompass complex activity-based models.* It has been mentioned in Chapter 3 that dynamic congestion games could easily model the possibility of short stops on a travel, or even longer activities as far as they are of constant duration. It would be of clear interest to be able to represent stops of variable time, the duration being a user's decision variable. This would require some extra research, and more specifically to reformulate the dynamic network loading problem.

- *The extension of the user-coordination algorithm to networks with tolls.* This latter has demonstrated a greater efficiency than the convex combination ones, but it is not clear how it could deal with generalized cost-based equilibrium rather than travel time-based ones.

# Bibliography


Adamo, V., Astarita, V., Florian, M., Mahut, M. and Wu, J. (1999), Modelling the spill-back of congestion in link based dynamic network loading models: a simulation model with application, *in* 'Proceedings of the 14th International Symposium on Transportation and Traffic Theory', pp. 555–573.

Aguiléra, V. (2010), Ladta user manual. Unpublished technical documentation.

Aguiléra, V. and Leurent, F. (2009), 'Large problems of dynamic network assignment and traffic equilibrium', *Transportation Research Record* **2**132, 122–132.

Aguiléra, V. and Wagner, N. (2009), A departure time choice model for dynamic assignment on interurban networks, *in* 'European Transport Conference (CD Rom edition)'.

Akamatsu, T. (2001), 'An Efficient Algorithm for Dynamic Traffic Equilibrium Assignment with Queues', *Transportation science* **3**5(4), 389–404.

Algers, S., Bernauer, E., Boero, M., Breheret, L., Di Taranto, C., Dougherty, M., Fox, K. and Gabard, J. (1997), 'Review of micro-simulation models', *Smartest Project deliverable D* **3**.

Anderson, S. and de Palma, A. (2004), 'The economics of pricing parking', *Journal of Urban Economics* **5**5(1), 1–20.

Arnott, R., de Palma, A. and Lindsey, R. (1991), 'Does providing information to drivers reduce traffic congestion?', *Transportation Research Part A: General* **2**5(5), 309–318.

Arnott, R., de Palma, A. and Lindsey, R. (1993), 'A structural model of peak period congestion: A traffic bottleneck with elastic demand', *American Economic Review* **8**3, 161–179.

Arnott, R., De Palma, A. and Lindsey, R. (1998), Recent developments in bottleneck model, *in* 'Road pricing, traffic congestion and the environment', Elgar's economics, pp. 70–110.

Becker, G. (1965), 'A Theory of the Allocation of Time', *The economic journal* pp. 493–517.

Beckmann, M., McGuire, C. and Winston, C. (1956), 'Studies in the economics of transportation', *Yale University Press* .

Bellei, G., Gentile, G. and Papola, N. (2005), 'A within-day dynamic traffic assignment model for urban road networks', *Transportation Research Part B: Methodological* **3**9(1), 1–29.

Ben-Akiva, M. and Lerman, S. (1985), *Discrete choice analysis: theory and application to travel demand*, The MIT Press.

Billingsley, P. (1995), *Probability and Measure*, John Wiley.

Bliemer, M. (2001), Analytical dynamic traffic assignment with interaction user-classes: theoretical advances and applications using a variational inequality approach, PhD thesis, Delft University of Technology, The Netherlands.

Bliemer, M. and Bovy, P. (2003), 'Quasi-variational inequality formulation of the multiclass dynamic traffic assignment problem', *Transportation Research Part B: Methodological* **37**(6), 501–519.

Brooke, G., Kenderick, D., Meeraus, A. and Release, G. (1996), '2.25 User's Guide', *GAMS Development Corporation, Washington* .

Cai, X., Kloks, T. and Wong, C. (1997), 'Time-varying shortest path problems with constraints', *Networks* **2**9(3), 141–150.

Cai, X., Sha, D. and Wong, C. K. (2007), *Time-Varying Network Optimization*, Springer Verlag, chapter Time-Varying Shortest Path Problems.

Chabini, I. (1998), 'Discrete dynamic shortest path problems in transportation applications: Complexity and algorithms with optimal run time', *Transportation Research Record* **1**645, 170–175.

Chabini, I. (2001), 'Analytical dynamic network loading problem: Formulation, solution algorithms, and computer implementations', *Transportation Research Record: Journal of the Transportation Research Board* **1**771(-1), 191–200.

Chabini, I. and Ganugapati, S. (2002), 'Parallel algorithms for dynamic shortest path problems', *International Transactions in Operational Research* **9**(3), 279–302.

Chandler, R., Herman, R. and Montroll, E. (1958), 'Traffic dynamics: studies in car following', *Operations Research* **6**(2), 165–184.

Chen, H.-K. and Hsueh, C.-F. (1998), 'A model and an algorithm for the dynamic user-optimal route choice problem', *Transportation Research Part B: Methodological* **3**2(3), 219–234.

Cooke, K. and Halsey, E. (1966), 'The shortest route through a network with time–dependent internodal transit times', *Journal of Mathematical Analysis and Applications* **1**2, 493–498.

Daganzo, C. (1985), 'The uniqueness of a time-dependent equilibrium distribution of arrivals at a single bottleneck', *Transportation Science* **1**9, 29–38.

Daganzo, C. (1994), 'The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory', *Transportation Research Part B: Methodological* **2**8, 269–287.

Daganzo, C. (1995), 'Properties of link travel times functions under dynamic loads', *Transportation Research Part B* **2**9, 95–98.

Daniele, P., Maugeri, A. and Oettli, W. (1998), 'Variational inequalities and time-dependent traffic equilibria', *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics* **3**26(9), 1059–1062.

De Palma, A. and Lindsey, R. (2002), 'Private roads, competition, and incentives to adopt time-based congestion tolling', *Journal of Urban Economics* **5**2(2), 217–241.

De Palma, A. and Marchal, F. (2002), 'Real cases applications of the fully dynamic METROPOLIS tool-box: an advocacy for large-scale mesoscopic transportation systems', *Networks and Spatial Economics* **2**, 347–369.

Dean, B. (1999), Continuous-time dynamic shortest path algorithms, Master's thesis, MIT.

Dean, B. (2004*a*), 'Algorithms for minimum-cost paths in time-dependent networks with waiting policies', *Networks* **4**4, 41–46.

Dean, B. (2004*b*), Shortest paths in FIFO time-dependent networks: Theory and algorithms, Technical report, MIT Department of Computer Science.

Deo, N. and Pang, C. (1984), 'Shortest path algorithms: a taxonomy and annotation', *Networks* **1**4, 275–323.

DeSerpa, A. (1971), 'A theory of the economics of time', *The Economic Journal* pp. 828–846.

Desrosiers, J., Soumis, F. and Desrochers, M. (1984), 'Routing with time windows by column generation', *Networks* **1**4, 545–565.

Dolecki, S., Salinetti, G. and Roger, J. (1983), 'Convergence of functions: equi-semicontinuity', *Transactions of the American Mathematical Society* pp. 409–429.

Dreyfus, S. (1969), 'An appraisal of some shortest-path algorithms', *Operation Research* **1**7, 395–412.

Dugundji, J. (1951), 'An extension of Tietze's theorem', *Pacific J. Math* **1**(3), 353–367.

Durlin, T. and Henn, V. (2008), Modélisation dynamique de l'affectation du trafic avec suivi explicite des ondes, *in* 'Proceedings of the INRETS-ENPC monthly seminar on Traffic modelling'.

Evans, A. W. (1972), 'On the theory of the valuation and allocation of time', *Scottish Journal of Political Economy* **1**9, 1–17.

Friesz, T., Bernstein, D., Smith, M., Tobin, R. and Wie, R. (1993), 'A variational inequality formulation of the dynamic network user equilibrium problem', *Operations Research* **4**1, 179–191.

Friesz, T. L., Luque, J., Tobin, R. and Wie, B. (2003), 'Dynamic network traffic assignment considered as a continuous time optimal control problem', *Operations research* **3**7, 893–901.

Gentile, G., Meschini, L. and Papola, N. (2007*a*), 'Spillback congestion in dynamic traffic assignment: a macroscopic flow model with time-varying bottlenecks', *Transportation Research B* pp. 1114–1138.

Gentile, G., Meschini, L. and Papola, N. (2007*b*), 'Spillback congestion in dynamic traffic assignment: A macroscopic flow model with time-varying bottlenecks', *Transportation Research Part B: Methodological* **4**1, 1114–1138.

Hart, S., Hildenbrand, W. and Kohlberg, E. (1974), 'On equilibrium allocation as distributions on the commodity space', *Journal of Mathematics economics* **1**, 159–167.

Hendrickson, C. and Kocur, G. (1981), 'Schedule delay and departure time choice in a determinist model', *Transportation Science* **1**5, 62–77.

Heydecker, B. G. and Addison, J. D. (2005), 'Analysis of dynamic traffic equilibrium with departure time choice', *Transportation Science* **3**9(1), 39–57.

Hildenbrand, W. (1974), *Core and equilibria of large economy*, Princeton University Press.

Hoogendoorn, S. and Bovy, P. (2000), 'Continuum modeling of multiclass traffic flow', *Transportation Research Part B: Methodological* **3**4(2), 123–146.

Hu, T.-Y. and Mahmassani, H. S. (1997), 'Day-to-day evolution of network flows under real-time information and reactive signal control', *Transportation Research Part C: Emerging Technologies* **5**(1), 51–69.

Janson, B. (1991), 'Dynamic traffic assignment for urban road networks', *Transportation Research Part B: Methodological* **2**5(2-3), 143–161.

Jayakrishnan, R., Mahmassani, H. S. and Hu, T.-Y. (1994), 'An evaluation tool for advanced traffic information and management systems in urban networks', *Transportation Research Part C: Emerging Technologies* **2**(3), 129–147.

Kaufman, D. and Smith, R. (1993), 'Fastest paths in time-dependent networks for intelligent vehicle-highway systems application', *Journal of Intelligent Transportation Systems* **1**, 1–11.

Khan, M. A. (1989), 'On Cournot-Nash equilibrium distributions for games with a nonmetrizable action space and upper semicontinuous payoffs', *Transactions of the American Mathematical Society* **3**15(1), 127–146.

Kuwahara, M. and Akamatsu, T. (1993), 'Dynamic equilibrium assignment with queues for a one-to-many OD pattern', *Transportation and Traffic Theory* **1**2, 185–204.

Lam, T. and Small, K. (2001), 'The value of time and reliability: measurement from a value pricing experiment', *Transportation Research Part E: Logistics and Transportation Review* **3**7(2-3), 231–251.

Lebacque, J. (1996), The Godunov scheme and what it means for first order traffic flow models, *in* 'Internaional symposium on transportation and traffic theory', pp. 647–677.

Leurent, F. (2003*a*), The multi-class flowing problem in a dynamic assignment model, *in* 'European Transport Conference (CD Rom edition)'.

Leurent, F. (2003*b*), On network assignment and supply demand equilibrium: an analysis framework and a simple dynamic model, *in* 'European Transport Conference (CD Rom edition)'.

Leurent, F. (2004), Le réseau dynamique : théorie et applications, *in* 'Proceedings of the INRETS-ENPC monthly seminar on Traffic modelling'.

Leurent, F. (2006), *Structures de réseau et modèles de cheminement*, Éd. Tec & doc.

Leurent, F. (2010), Les contraintes de capacité en transport public de voyageurs: une analyse systémique. Working paper, english version submitted to the European Transport Research Review.

Leurent, F., Aguiléra, V. and Mai, H. (2007), Convergence criteria and computational efficiency of a dynamic traffic assignment model., *in* 'World Conference on Transport Research(WCTR), first CD Rom edition'.

Leurent, F., Mai, H. and Aguilèra, V. (2006), Sur la caractérisation d'un équilibre dy,namique du trafic: formulations analaytiques, mesurage de convergence, algorithmes d'équilibrage, *in* 'Seminaire INRETS-ENPC de modélisation du trafic'.

Leurent, F. and Wagner, N. (2009), 'User Equilibrium in a Bottleneck Under Multipeak Distribution of Preferred Arrival Time', *Transportation Research Record: Journal of the Transportation Research Board* **2**091, 31–39.

Lighthill, M. and Whitman, J. (1955), 'On kinematic waves, ii: A theory of traffic flow on long crowded roads', *Proceeding of Royal Society* **A**229, 281–345.

Lindsey, C. R. and Verhoef, E. T. (1999), Congestion modelling, Tinbergen institute discussion papers, Tinbergen Institute.

Lindsey, R. (2004), 'Existence, uniqueness and trip cost function properties of user equilibrium in the bottleneck model with multiple user classes', *Transportation science* **3**8, 293–314.

Lo, H. K. and Szeto, W. Y. (2002), 'A cell-based variational inequality formulation of the dynamic user optimal assignment problem', *Transportation Research Part B: Methodological* **3**6(5), 421–443.

Lusin, N. (1912), 'Sur les propriétés des fonctions mesurables', *Comptes Rendus Acad. Sci. Paris* **1**54, 1688–1690.

Mahmassani, H. S., Hu, T. and Jayakrishnan, R. (1995), Dynamic traffic assignment and simulation for advanced network informatics (DYNASMART), *in* Springer, ed., 'Urban traffic networks: Dynamic flow modeling and control'.

Mas-Colell, A. (1984), 'On a theorem of Schmeidler', *Journal of Mathematical Economics* **1**3(3), 201–206.

May, A. D. and Keller, H. M. (1967), 'A deterministic queuing model', *Transportation research* **1**, 117–128.

Merchant, D. and Nemhauser, G. L. (1978), 'A model and an algorithm for the dynamical traffic assignment problems', *Transportation Research* **1**2, 183–199.

Meunier, F. and Wagner, N. (2010), Equilibrium results for dynamic congestion games. To be published in the November 2010 issue of Transportation Science.

Milchtaich, I. (2005), 'Topological conditions for uniqueness of equilibrium in networks', *Math. of Op. Res.* **3**0, 226–244.

Mounce, R. (2003), Convergence in a continuous dynamic traffic assignment model, PhD thesis, University of York, UK.

Mounce, R. (2006), 'Convergence in a continuous dynamic queuing model for traffic networks', *Transportation Research Part B: Methodological* **4**0, 779–791.

Mounce, R. (2007), 'Existence of equilibrium in a continuous dynamic queueing model for traffic networks', *Proceedings of the 4th IMA Conference on Mathematics in Transport* .

Mounce, R. and Carey, M. (2010), 'Route swapping in dynamic traffic networks', *Transportation Research Part B: Methodological* **I**n Press, Corrected Proof, –.

Nguyen, S., Pallottino, S. and Malucelli, F. (2001), 'A modeling framework for passenger assignment on a transport network with timetables', *Transportation Science* **3**5(3), 238–249.

Orda, A. and Rom, R. (1990), 'Shortest-path and minimum-delay algorithms in networks with time-dependent edge-length', *Journal of the ACM (JACM)* **3**7, 607–625.

Orda, A. and Rom, R. (1991), 'Minimum weight paths in time-dependent networks', *Networks* **2**1(3), 295–319.

Papon, F. (1992), Les routes de première classe : un péage urbain choisi par l'usager, *in* 'La mobilité urbaine : de la paralysie au péage', Edition du P.P.S.H., recherches en Sciences Humaines, pp. 143–164.

Patriksson, M. (1999), *Nonlinear programming and variational inequality problems: a unified approach*, Kluwer Academic Pub.

Payne, H. (1971), 'Models of freeway traffic and control', *Mathematical models of public systems* **1**(1), 51–61.

Ran, B. and Boyce, D. E. (1996), *Modeling dynamic transportation networks: an intelligent transportation system oriented approach*, Springer Verlag.

Richards, P. (1956), 'Shock waves on the highway', *Operations Research* **4**, 42–51.

Rosenthal, R. (1973), 'A class of games possessing pure-strategy nash equilibria', *International Journal of Game Theory* **2**, 65–67.

Rubio-Ardanaz, J., Wu, J. and Florian, M. (2003), 'Two improved numerical algorithms for the continuous dynamic network loading problem', *Transportation research Part B: Methodological* **37**, 171–190.

Rudin, W. (2009), *Analyse réelle et complexe: cours et exercices*, Dunod.

Sautter, A. (2007), Accessibilité et Modulation tarifaire, Technical report, Cofiroute.

Schmeidler, D. (1973), 'Equilibrium points on non-atomic games', *J. Satistic. Phys* **7**, 295–300.

Scilab Consortium (2010), 'Scilab, the open source platform for numerical computation', `http://www.scilab.org/`.

Sheffi, Y. (1985), *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice-Hall.

Shoup, D. (1997), 'The high cost of free parking', *Journal of Planning Education and Research* **17**(1), 3.

Small, K. (1982), 'The scheduling of consumer activities: work trips', *The American Economic Review* **72**(3), 467–479.

Small, K. and Yan, J. (2001), 'The value of" value pricing" of roads: Second-best pricing and product differentiation', *Journal of Urban Economics* **4**9(2), 310–336.

Smith, M. (1984), 'The existence of a time-dependent equilibrium distribution at arrivals at a single bottleneck', *Transportation Science* **1**8, 385–394.

Smith, M., Duncan, G. and Druitt, S. (1995), PARAMICS: microscopic traffic simulation for congestion management, *in* 'IEE Seminar Digests', Vol. 8.

Smith, M. and Wisten, M. (1995), 'A continuous day-to-day traffic assignment model and the existence of a continuous dynamic user equilibrium', *Annals of Operations Research* **6**0, 59–79.

Szeto, W. Y. (2008), 'Enhanced lagged cell-transmission model for dynamic traffic assignment', *Transportation Research Record* **2**085, 76–85.

Szeto, W. Y. and Lo, H. K. (2004), 'A cell-based simultaneous route and departure time choice model with elastic demand', *Transportation Research Part B: Methodological* **3**8(7), 593–612.

Tong, C. O. and Wong, S. C. (2000), 'A predictive dynamic traffic assignment model in congested capacity-constrained road networks', *Transportation Research* pp. 625–644.

Tong, C. and Wong, S. (1999), 'A schedule-based time-dependent trip assignment model for transit networks', *Journal of Advanced Transportation* **3**3(3), 371–388.

Tong, C., Wong, S., Poon, M. and Tan, M. (2001), 'A schedule-based dynamic transit network model-Recent advances and prospective future research', *Journal of Advanced Transportation* **3**5(2), 175–195.

Topsoe, F. (1970), *Topology and measure*, Vol. 133 of *Lecture Notes in Math.*, Springer-Verlag, New York.

van den Berg, V. and Verhoef, E. T. (2010), 'Congestion tolling in the bottleneck model with heterogeneous values of time', *Transportation Research Part B: Methodological* **I**n Press, Corrected Proof.

Van der Zijpp, N. and Koolstra, K. (2002), 'Multiclass continuous-time equilibrium model for departure time choice on single-bottleneck network', *Transportation Research Record* **1**783, 134–141.

Verhoef, E., Nijkamp, P. and Rietveld, P. (1996), 'Second-Best Congestion Pricing: The Case of an Untolled Alternative', *Journal of Urban Economics* **4**0(3), 279–302.

Verhoef, E. T. and Small, K. A. (1999), Product differentiation on roads: Second-best congestion pricing with heterogeneity under public and private ownership, ERSA conference papers, European Regional Science Association.

Vickrey, W. (1969), 'Congestion theory and transport investment', *American Economic Review* **5**9, 251–261.

Wardrop, J. (1952), 'Some theoretical aspects of road traffic research', *Proceeding of the Institution of Civil Engineers II* pp. 325–378.

Wie, B., Tobin, R. and Carey, M. (2002), 'The existence, uniqueness and computation of an arc-based dynamic network user equilibrium formulation', *Transportation Research Part B: Methodological* **3**6(10), 897–918.

Wu, J., Chen, Y. and Florian, M. (1998), 'The continuous dynamic network loading problem: a mathematical formulation and solution method', *Transportation research Part B: Methodological* **3**2B, 173–187.

Xu, Y., Wu, J., Florian, M. and Zhu, D. (1999), 'Advances in the continuous dynamic network loading problem', *Transportation science* **3**3, 341–353.

Zhu, D. and Marcotte, P. (2000), 'On the existence of solutions to the dynamic user equilibrium problem', *Transportation Science* **3**4(4), 402–414.

Ziliaskopoulos, A., Kotzinos, D. and Mahmassani, H. S. (1997), 'Design and implementation of parallel time-dependent least time path algorithms for intelligent transportation systems applications', *Transportation Research Part C: Emerging Technologies* **5**(2), 95–107.

Ziliaskopoulos, A. and Mahmassani, H. (1993), 'Time-dependent, shortest-path algorithm for real-time intelligent vehicle highway system applications', **1**408, 94–104.

Ziliaskopoulos, A. and Wardell, W. (2000), 'An intermodal optimum path algorithm for dynamic multimodal networks', *European Journal of Operational Research* **1**25, 468–502.

# Appendices

Appendix A

# Proof of Proposition 5.6 of Chapter 5

**Proposition 5.6.** *$W_m(h_0) := \min_i \hat{\tau}_i(h_0)$ is a continuous and decreasing function.*

*Proof of Proposition 5.6.* Consider an interval $[h_m, h_M]$ included in an off-peak period and denote $P_i = [p_{i-1}, p_i]$, $i = 1, \ldots, 2n$, the sequence of peak and off-peak periods after $h_M$. The proof proceeds in three steps. We shall first define for each $i$ a function $h_0 \mapsto \hat{\hbar}(h_0)$ on $[h_m, h_M]$ that takes its value in $P_i$. Second, some properties of these functions will be established. Third, we shall conclude about $\min_i \hat{\tau}_i$. We shall make use of an auxiliary function defined as follows:

$$(h_0, \bar{h}) \mapsto \Delta(h_0, \bar{h}) := k.(\bar{h} - h_0) - X_p(\bar{h}) + X_p(h_0))$$

**Step I: Defining $\hat{\hbar}_i(h_0)$.** For any $h_0$ in $[h_m, h_M]$ let us define $\hat{\hbar}_0, \ldots, \hat{\hbar}_{2n}$ by setting $\hat{\hbar}_0 := h_0$ and by using the following recursive rule. For any $i$ from 1 to $2n$, try to solve the equation $\Delta(h_0, \bar{h}) = 0$ in $\bar{h}$ on $P_i$: if there is a solution $\bar{h}$ then set $\hat{\hbar}_i$ to $\bar{h}$ , else set $\hat{\hbar}_i$ to either $p_i$ or $p_{i-1}$ according to the following table of cases.

| Case | $\Delta > 0$ on $P_i$ | $\Delta < 0$ on $P_i$ |
|---|---|---|
| $i$ odd | $p_i$ | $p_{i-1}$ |
| $i$ even | $p_{i-1}$ | $p_i$ |

Table 1.1: Table of cases for the prolongation of $\hat{\hbar}_i$

The derivation of a sequence $(\hat{\hbar}_i)$, illustrated in Figure A.1, stands as an *ad-hoc* extension of formula (5.14) so as to address degeneracy in the number

of early and late sub-periods. When several neighboring peak periods give rise to a common queued period, then there might be less actual early and late sub-periods than peak periods. The proposed extension deals with this issue by adding "fake" subperiods of null size.

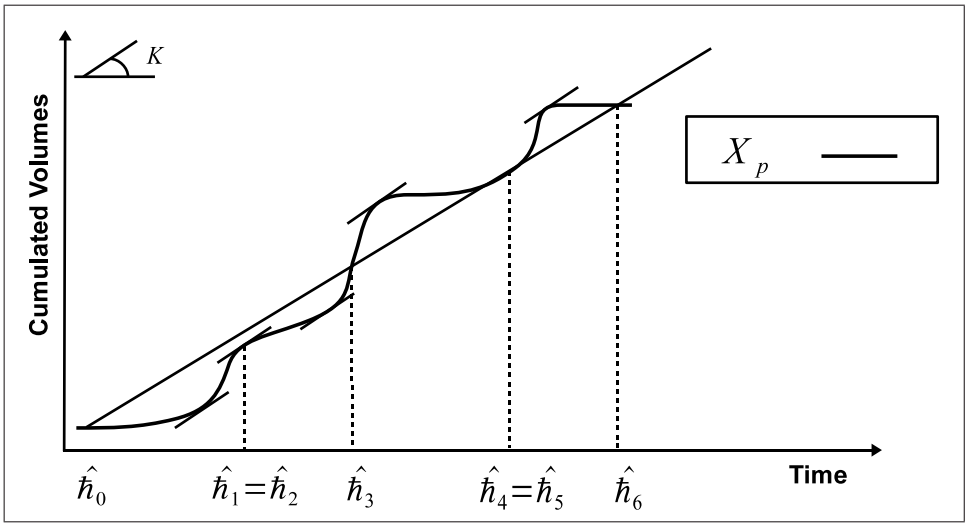



Figure A.1: Derivation of would-be critical arrival times

**Step II: Properties of $\hat{h}_i(h_0)$.** Let us show that the functions $\hat{h}_i(.)$ are continuous and monotonic, decreasing if $i$ is odd, meaning an off-peak period $P_i$, or increasing if $i$ is even meaning a peak period $P_i$. In the case of even $i$, consider $\Delta$ on $]h_m, h_M[\times]p_{i-1}, p_i[$. This is a continuous function with partial derivatives with respect to $\bar{h}$ and $h_0$ as follows:

$$\Delta_{h_0}(h_0, \bar{h}) = x_p(h) - k < 0 \text{ and } \Delta_{\bar{h}}(h_0, \bar{h}) = k - x_p(\bar{h}) > 0.$$

Six cases can arise, all them depicted in Figure A.2.

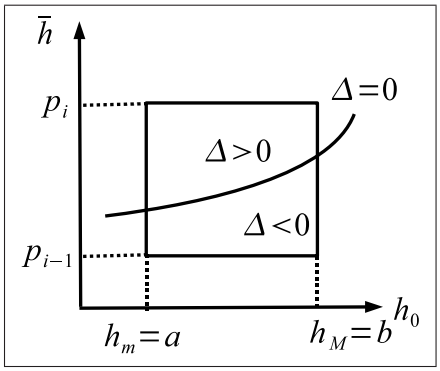
- Cases 5 and 6 are degenerated situations where $\Delta(h, \bar{h}) \neq 0$ and where consequently $\hat{h}_i(h_0) = p_i$ (case 5) or $\hat{h}_i(h) = p_{i-1}$ (case 6) for all $h_0 \in [h_m, h_M]$.

- In cases 1 to 4, Equation $\Delta(h, \bar{h}) = 0$ defines implicitly a function $\hat{h}_i(h)$ on an interval $]a, b[$ in such a way that $(a, \lim_a \hat{h}_i)$ and $(b, \lim_b \hat{h}_i)$ lie on the boundary of $]h_m, h_M[\times]p_{i-1}, p_i[$. Hence, $a$ that $a = h_m$ or $\lim_a \hat{h}_i = p_{i-1}$ and $b$ is such that $b = h_M$ or $\lim_b \hat{h}_i = p_i$. Furthermore, for all $\bar{h}$ we have $\Delta_{\bar{h}}(h_0, \bar{h}) < 0$ for $h < a$ and $\Delta_{\bar{h}}(h_0, \bar{h}) > 0$ for $h > b$. Prolongating each $\hat{h}_i$ on $[h_m, h_M]$ by the process defined above is thus continuous.

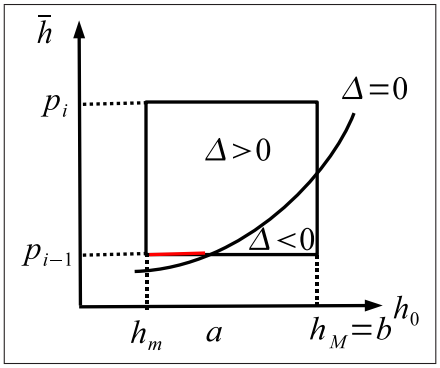
In all 6 cases, $\bar{h}_i$ is a continuous functions. In cases 5 and 6, it is trivially

increasing as a constant function. In cases 1 to 4, $\frac{d\hat{h}_i}{dh} = -\frac{\Delta_{h_0}(h_0, \bar{h})}{\Delta_{\bar{h}}(h_0, \bar{h})} > 0$ on $[a, b]$ (implicit function theorem) and $\bar{h}_i$ is constant elsewhere.

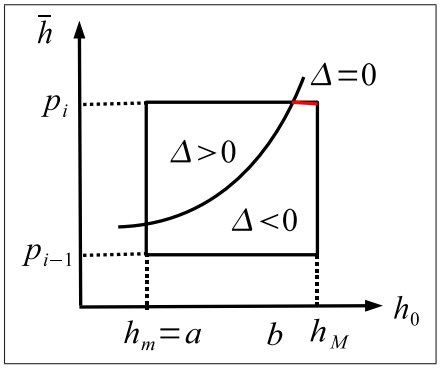
The case when $i$ is odd is similar.



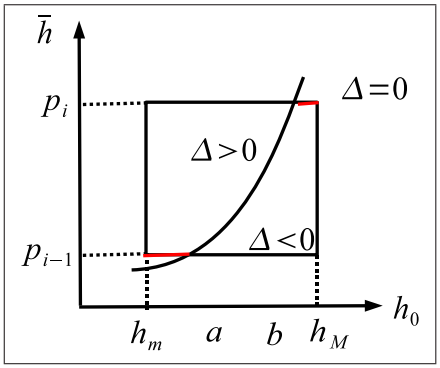

Case 1:
$a = h_m$ and $b = h_M$



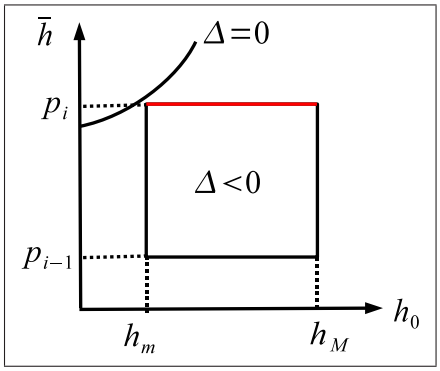

Case 2:
$\lim_a \hat{\hbar}_i = p_{i-1}$ and $b = h_M$



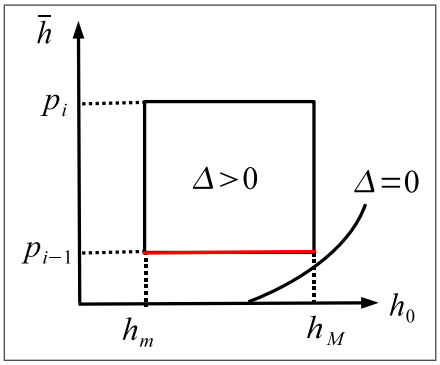

Case 3:
$a = h_m$ and $\lim_b \hat{\hbar}_i = p_i$




Case 4
$\lim_a \hat{\hbar}_i = p_{i-1}$ and $\lim_{h\to b} \hat{\hbar}_i = p_i$




Case 5




Case 6

Figure A.2: Possible cases for implicit equation $\Delta(h, \bar{h}) = 0$ for a peak period $P_i$ The red line depicts the prolongation of $\hat{\hbar}_i$ using the rules of Table 1.1

**Step III: Proof that $W_m$ is a continuous and decreasing function.** For each $h_0$, derive $\hat{h}_i(h_0)$ and $\hat{\tau}_i(h_0)$ from $\hat{\hbar}_i(h_0)$ on the basis of (5.16) and (5.17). By straightforward substitution of (5.16) into (5.17) we get that

$$\hat{\tau}_{i+1}(h_0) = \hat{\tau}_i(h_0) + \frac{x_i^+ - k}{x_i^+}(\hat{\hbar}_{i+1}(h_0) - \hat{\hbar}_{i-1}(h_0)) \tag{1.1}$$

where $x_i^+$ is defined as in (5.16). As $\hat{\hbar}_{i+1} - \hat{\hbar}_i$ is a decreasing [resp. increasing] with respect to $h_0$ if $i$ is even [resp. odd] hence $x_i^+$ is positive [resp. negative] the incremental part in (1.1) is a decreasing function of $h_0$. Then each $\hat{\tau}_i$ is a decreasing function of $h_0$, owing to recursion and to the initial condition $\hat{\tau}_0 = 0$. Concluding, the minimum $W_m := \min_i \hat{\tau}_i$ is a continuous and decreasing function of $h_0$ as the minimum of a sequence of such functions.

□

Appendix B

# Proofs of Chapter 6

**Lemma 6.15.** *Assuming a DUE of type a, $\nu_1^\star$ and $\nu_2^\star$ are characterized by the following relationships:*

$$\tilde{g}(\nu_1^\star) = \eta V(\nu_1^\star) + \nu_1^\star t + \rho p$$

*and*

$$\eta V(\nu_2^\star) = (1-\rho).p$$

*Proof of Lemma 6.15.*

**First Equality** As a solution to the DUE problem, $\tau_1$, $h_-^1$, $h_+^1$ and $\nu_1^\star$ are such that:

$$\tau_1\big(h_-^1(\nu_1^\star)\big) = \tau_1\big(h_+^1(\nu_1^\star)\big) = 0$$

Then, from Equation (6.22), it comes:

$$\tau_2\big(h_-^2(\nu_1^\star)\big) = \tau_2\big(h_+^2(\nu_1^\star)\big) = t$$

As $\tilde{g} := \min_i \tilde{g}_i$ satisfies Equation (6.15):

$$\tilde{g}(\nu_1^\star) = -\alpha h_-^1(\nu_1^\star) + \nu_1^\star.t = \beta h_+^1(\nu_1^\star) + \nu_1^\star.t$$

and

$$\tilde{g}(\nu_1^\star) = -\alpha h_-^2(\nu_1^\star) + \nu_1^\star.t + p = \beta h_+^2(\nu_1^\star) + \nu_1^\star.t + p$$

Which, combined with (6.16), yields:

$$\begin{aligned}\tilde{g}(\nu_1^\star) &= \eta/k.\left[\rho.\big(h_+^2(\nu_1^\star) - h_-^2(\nu_1^\star)\big) + (1-\rho).\big(h_+^1(\nu_1^\star) - h_-^1(\nu_1^\star)\big)\right] + \nu_1^\star t + \rho p \\ &= \eta V(\nu_1^\star) + \nu_1^\star t + (1-\rho)p\end{aligned}$$

**Second equality** As a solution to the DUE problem, $h_-^2$, $h_+^2$ and $\nu_2^\star$ are such that: $h_+^2(\nu_2^\star) = h_-^2(\nu_2^\star) = 0$. According to Equation (6.15):

$$\tilde{g}(\nu_2^\star) = \alpha h_-^1(\nu_2^\star) + \nu_2^\star.t + \nu_2^\star.\tau(h_-^1(\nu_2^\star)) = -\beta h_+^1(\nu_2^\star) + \nu_2^\star.t + \nu_2^\star.\tau(h_+^1(\nu_2^\star))$$

and

$$\tilde{g}(\nu_2^\star) = \nu_2^\star \tau(h_-^1(\nu_2^\star)) + p = \nu_2^\star \tau(h_+^1(\nu_2^\star)) + p$$

The first equation gives:

$$\tilde{g}(\nu_2^\star) = \eta k.\big(h_+^1(\nu_2^\star) - h_-^1(\nu_2^\star)\big) + \nu_2^\star.t + \nu_2^\star.\tau(h_+^1(\nu_2^\star))$$

and combining it with (6.16):

$$\tilde{g}(\nu_2^\star) = \frac{\eta}{1-\rho} V(\nu_2^\star) + \nu_1^\star.\big(t + \tau(h_+^1((\nu_2^\star))\big)$$

According to Equation (6.22), $\tau(h_+^2(\nu_2^\star)) = \tau(h_+^1(\nu_2^\star)) + t$, so it comes:

$$\eta V(\nu_2^\star) = (1-\rho)p$$

□

**Lemma 6.16.** *The boundary conditions are :*

$$\frac{\partial \tilde{g}}{\partial \nu}(\nu_M) = 0$$

$$\tilde{g}(\nu_M) = \frac{\eta}{\rho}\Big(V(\nu_M) - V(\nu_1^\star)(1-\rho)\Big) + \nu_1^\star t + \rho p$$

*Proof of Lemma 6.16.* As $\tau_2$ and $h_-^2$ are solution to the DUE2R problem,

$$\tau_2(h_-^2(\nu_M)) = 0$$

According to Equation (6.15),

$$\tilde{g}(\nu_M) = \alpha h_+^2(\nu_M) + p = -\beta h_+^2(\nu_M) + p$$

Hence,

$$\tilde{g}(\nu_M) = \eta k.\big(h_-^2(\nu_M) - h_+^2(\nu_M)\big) + p$$

Using Lemma 6.15, one can get

$$\eta k\big(h_-^2(\nu_1^\star) - h_+^2(\nu_1^\star)\big) = V(\nu_1^\star) + (\rho - 1)p$$

and Equation (6.16) leads to

$$h_-^2(\nu_M) - h_+^2(\nu_M) = h_-^2(\nu_1^\star) - h_+^2(\nu_1^\star) + \frac{V(\nu_M) - V(\nu_1^\star)}{\rho k}$$

Combining the two last equations gives

$$h_-^2(\nu_M) - h_+^2(\nu_M) = \frac{\rho - 1}{\rho k} V(\nu_1^\star) + \frac{1}{\rho k} V(\nu_M) + \frac{\rho - 1}{\eta k} p$$

So finally it comes

$$\tilde{g}(\nu_M) = \frac{\eta}{\rho}\Big(V(\nu_M) - V(\nu_1^\star)(1 - \rho)\Big) + \rho p$$

□

**Proposition 6.11** (Equivalency of the DUE2R and the two ECP2R)**.**

*(1) If the two quadruples* $\Theta_i = (h_-^i, h_+^i, \tau_i, \nu_i^\star)$*, for* $i \in \{1; 2\}$*, solves the DUE2R problem, then* $\tilde{g} := \min_i \tilde{g}_i(.; \Theta_i)$ *solves either the ECP2Ra (and then* $\nu_1^\star \neq \nu_2^\star$*) or the ECP2Rb (and then* $\nu_1^\star = \nu_2^\star$*).*

*(2) Consider* $(\tilde{g}_a, \nu_1^\star, \nu_2^\star)$ *and* $(\tilde{g}_b, \nu^\star)$ *the respective solutions of the ECP2Ra and the ECP2Rb. Then:*

- *(i)* $\nu_1^\star = \nu_2^\star \quad \Rightarrow \nu^\star = \nu_1^\star = \nu_2^\star$ *and* $\tilde{g}_a = \tilde{g}_b$ *are solutions to the DUE2R problem.*
- *(ii)* $\nu_1^\star < \nu_2^\star \quad \Rightarrow (\tilde{g}_a, \nu_1^\star, \nu_2^\star)$ *is a solution of the DUE2R problem.*
- *(iii)* $\nu_1^\star > \nu_2^\star \quad \Rightarrow (\tilde{g}_b, \nu^\star, \nu^\star)$ *is a solution of the DUE2R problem.*

*Proof of the "sufficiency" part of Proposition 6.11.*
**(ii).** Consider $\tilde{g}$ the solution to the ECP2Ra and let $(\Theta_i)_{i\in\{1,2\}} = (h^i_-, h^i_+, \nu^i_\star)_{i\in\{1,2\}}$ and $(\tau)_{i\in\{1,2\}}$ be defined as in Definition 6.10. Proceeding as in Proposition 6.3, it can easily be proven that the triple $(h^1_-, h^1_+, \tau_1)$ is a solution to the DUE problem with one route of capacity $\rho k$ and a VoT distribution of $N_1(.;\Theta_1)$. Similarly, the triple $(h^2_-, h^2_+, \tau_2)$ is a solution to the DUE problem with one route of capacity $(1-\rho)k$ and a VoT distribution of $N_2(.;\Theta_2)$.

It remains to show that Equations (6.15) and (6.16) hold. Consider $\tilde{g}_1(\nu;\theta_1,\tau_1)$ on $[\nu_m,\nu^\star_1]$:

$$\begin{aligned}
\tilde{g}_1(\nu;\theta_1,\tau_1) &= \nu.\Big(\tau_1\big(h^1_-(\nu)\big)+t\Big) - \alpha.h^1_-(\nu) \\
&= \nu\frac{\partial\tilde{g}}{\partial\nu}(\nu) + \tilde{g}(\nu^\star_1) - \nu^\star_1.t - \int_\nu^{\nu^\star_1} \nu\frac{\partial^2\tilde{g}}{\partial\nu^2}(\nu)\,d\nu && \text{(Replace } \tau_1 \text{ and } h^1_- \text{ by} \\
&&& \text{their expr. in Def. 6.10)} \\
&= \tilde{g}(\nu^\star_1) - \int_{\nu^\star_1}^{\nu} \frac{\partial}{\partial\nu}\left(\nu\frac{\partial\tilde{g}}{\partial\nu}(\nu)\right)d\nu + \int_{\nu^\star_1}^{\nu}\nu\frac{\partial^2\tilde{g}}{\partial\nu^2}(\nu)\,d\nu && (\text{As } \nu^\star_1 t = \nu^\star_1\partial\tilde{g}/\partial\nu(\nu^\star_1)) \\
&= \tilde{g}(\nu^\star_1) + \int_{\nu^\star_1}^{\nu}\frac{\partial\tilde{g}}{\partial\nu}(\nu)\;d\nu \\
&= \tilde{g}(\nu)
\end{aligned}$$

Similarly, it can be shown that $\tilde{g}_2(\nu;\Theta_2,\tau_2) = \tilde{g}(\nu)$ on $[\nu^\star_2,\nu_m]$. Thus, Equation (6.15) is satisfied for $\nu\in[\nu^\star_2,\nu^\star_1]$.

Let us now prove it on $[\nu_m,\nu^\star_2]$. Recall that for any $\nu\in[\nu_m,\nu^\star_2]$:

$$h^2_-(\nu) = h^2_-(\nu^\star_2) \quad \text{(by definition of } h^2_-)$$

and that for all $h$:

$$\tilde{g}_1(\nu;\Theta_1,\tau_1) \le g_1(h;\nu,\tau_1) \quad \text{(as } (\Theta_1,\tau_1) \text{ is a DUE)}$$

Consequently:

$$\begin{aligned}
\tilde{g}_1(\nu;\Theta_1,\tau_1) \le\;& g_1(h^1_-(\nu^\star_2);\nu,\tau_1) && \text{(According to the Ineq. above)} \\
\le\;& \nu.\tau_1\big(h^1_-(\nu^\star_2)\big) + \nu.t - \alpha h^1_-(\nu^\star_2) && \text{(By definition of } \tilde{g}_1 \text{ )} \\
\le\;& \nu.\tau_2\big(h^2_-(\nu^\star_2)\big) - \alpha h^2_-(\nu^\star_2) && \text{(According to the definition of} \\
&&& \tau_1 \text{ and } \tau_2 \text{ in Def. 6.10)} \\
\le\;& \nu.\tau_2\big(h^2_-(\nu^\star_2)\big) - \alpha h^2_-(\nu^\star_2) + p && (\text{As } -\alpha h^2_-(\nu^\star_2) = -\alpha h^1_-(\nu^\star_2) + p) \\
\le\;& \tilde{g}_2(\nu;\Theta_2,\tau_2)
\end{aligned}$$

It remains to prove Equation (6.15) on $[\nu_2^\star, \nu_M]$. The proof uses similar arguments as on $[\nu_m, \nu_1^\star]$. Recall that for any $\nu \in [\nu_2^\star, \nu_M]$:

$$h_-^1(\nu) = h_-^1(\nu_1^\star) \quad \text{(by definition of } h_-^1\text{)}$$

and for all $h$:

$$\tilde{g}_2(\nu; \Theta_2, \tau_2) \leq g_2(h; \nu, \tau_2) \quad \text{(as } (\Theta_2, \tau_2) \text{ is a DUE on route 2)}$$

Moreover note that:

$$\begin{aligned} \tilde{g}_1(\nu_1^\star; \Theta_1, \tau_1) = \tilde{g}_2(\nu_1^\star; \Theta_2, \tau_2) \text{ and } \tau_1(\nu_1^\star) + t = \tau_1(\nu_2^\star) \\ \Rightarrow -\alpha h_-^2(\nu_1^\star) = -\alpha h_-^1(\nu_1^\star) + p \end{aligned}$$

Consequently:

$$\begin{array}{rcll} \tilde{g}_2(\nu; \Theta_2, \tau_2) \leq & g_2(h_-^2(\nu_1^\star); \nu, \tau_2) & \text{(According to the Ineq. above)} \\ \leq & \nu.\tau_2\big(h_-^2(\nu_1^\star)\big) - \alpha h_-^1(\nu_1^\star) + p & \text{(By definition of } \tilde{g}_2 \text{ )} \\ \leq & \nu.\tau_1\big(h_-^1(\nu_2^\star)\big) + \nu.t - \alpha h_-^1(\nu_2^\star) + p & \text{(According to the definition of} \\ & & \tau_1 \text{ and } \tau_2 \text{ in Def. 6.10)} \\ \leq & \nu.\tau_1\big(h_-^1(\nu_2^\star)\big) - \alpha h_-^1(\nu_2^\star) & \text{(As } -\alpha h_-^2(\nu_1^\star) = -\alpha h_-^1(\nu_1^\star) + p\text{)} \\ \leq & \tilde{g}_1(\nu; \Theta_1, \tau_1) & \end{array}$$

Finally, we are going to prove Equation (6.16).
For any $\nu \in [\nu_m, \nu_1^\star]$:

$$\begin{aligned} N_1(\nu; \Theta_1) &= (1-\rho)k.(h_+^1(\nu) - h_-^1(\nu)) \\ &= \frac{1-\rho}{\eta}\left(\tilde{g}(\nu_1^\star) - \nu_1^\star t - \int_{\nu_1^\star}^{\nu} \nu \frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu)\, d\nu\right) \end{aligned}$$

For any $\nu \in [\nu_2^\star, \nu_M]$:

$$\begin{aligned} N_2(\nu; \Theta_2) &= \rho k.(h_+^2(\nu) - h_-^2(\nu)) \\ &= \frac{\rho}{\eta}\left(\tilde{g}(\nu_M) - p - \int_{\nu_M}^{\nu} \nu \frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu)\, d\nu\right) \end{aligned}$$

By replacing $\tilde{g}(\nu_M)$ and $\tilde{g}(\nu_1^\star)$ by their expression, it is straightforward that for $\nu \in [\nu_2^\star, \nu_1^\star]$:

$$N_1(\nu; \Theta_1) + N_2(\nu; \Theta_2) = V(\nu)$$

Now recall that $h^1_+(\nu) = h^1_+(\nu^\star_1)$ and $h^1_-(\nu) = h^1_-(\nu^\star_1)$ for $\nu > \nu^\star_1$. Thus $N_1(\nu;\Theta_1) = N_1(\nu^\star_1;\Theta_1)$ for such $\nu$. Then, for $\nu \in [\nu^\star_1, \nu_M]$:

$$\begin{aligned}
&N_1(\nu;\Theta_1) + N_2(\nu;\Theta_2)\\
&= N_1(\nu^\star_1;\Theta_1) + N_2(\nu;\Theta_2)\\
&= \frac{1-\rho}{\eta}\left(\tilde{g}(\nu^\star_1) - \nu^\star_1 t\right) + \frac{\rho}{\eta}\left(\tilde{g}(\nu_M) - p - \int_{\nu_M}^{\nu} \nu \frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu)\, d\nu\right)\\
&= \frac{1-\rho}{\eta}\big(\eta V(\nu^\star_1) + \rho p\big) + \frac{\rho}{\eta}\big(\tilde{g}(\nu_M) - p\big) - \int_{\nu_M}^{\nu} v(\nu)\, d\nu\\
&= V(\nu)\\
&\quad \text{(Replace } \tilde{g}(\nu^\star_1) \text{ and then } \tilde{g}(\nu_M) \text{ by their expr. in Def. 6.8)}
\end{aligned}$$

Finally for $\nu \in [\nu_m, \nu^\star_2]$:

$$\begin{aligned}
&N_1(\nu;\Theta_1) + N_2(\nu;\Theta_2)\\
&= N_1(\nu;\Theta_1) + N_2(\nu^\star_2;\Theta_2)\\
&= N_1(\nu;\Theta_1) - N_1(\nu^\star_2;\Theta_1) + N_1(\nu^\star_2;\Theta_1) + N_2(\nu^\star_2;\Theta_2)\\
&= N_1(\nu;\Theta_1) - N_1(\nu^\star_2;\Theta_1) + V(\nu^\star_2)\\
&= \frac{1-\rho}{\eta}\int_{\nu^\star_2}^{\nu} \nu \frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu)\, d\nu + V(\nu^\star_2)\\
&= V(\nu)
\end{aligned}$$

**(iii).** Consider $\tilde{g}$ the solution to the ECP2Rb and let $(\Theta_i)_{i\in\{1,2\}} = (h^i_-, h^i_+, \nu^i_\star)_{i\in\{1,2\}}$ and $(\tau)_{i\in\{1,2\}}$ be defined as in Definition 6.10. As for the part (ii) of this proof, it is required to prove that Equations (6.15) and (6.16) hold.

The arguments used in part (ii) of this proof to show that Equation (6.15) is satisfied are also valid here, once $\nu^\star_1$ and $\nu^\star_2$ are replaced by $\nu^\star$. Thus it remains to prove Equation (6.16).
For any $\nu \in [\nu_m, \nu^\star]$:

$$\begin{aligned}
N_1(\nu;\Theta_1) &= (1-\rho)k.(h^1_+(\nu) - h^1_-(\nu))\\
&= \frac{1-\rho}{\eta}\left(\tilde{g}(\nu^\star_1) - \nu^\star t - \int_{\nu^\star}^{\nu} \nu \frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu)\, d\nu\right)
\end{aligned}$$

For any $\nu \in [\nu^\star, \nu_M]$:

$$\begin{aligned}
N_2(\nu;\Theta_2) &= \rho k.(h^2_+(\nu) - h^2_-(\nu))\\
&= \frac{\rho}{\eta}\left(\tilde{g}(\nu_M) - p - \int_{\nu_M}^{\nu} \nu \frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu)\, d\nu\right)
\end{aligned}$$

Consequently on $[\nu^\star, \nu_M]$:

$$\begin{aligned} & N_1(\nu;\Theta_1) + N_2(\nu;\Theta_2) \\ &= N_1(\nu^\star;\Theta_1) + N_2(\nu;\Theta_2) \\ &= \frac{1-\rho}{\eta}\left(\tilde{g}(\nu_1^\star) - \nu^\star t\right) + \frac{\rho}{\eta}\left(\tilde{g}(\nu_M) - p - \int_{\nu_M}^{\nu} \nu \frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu)\, d\nu\right) \\ &= V(\nu^\star) - V(\nu^\star) + V(\nu_M) + \frac{\rho}{\eta}\int_{\nu_M}^{\nu} \nu \frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu)\, d\nu \\ &= V(\nu) \end{aligned}$$

and on $[\nu_m, \nu^\star]$:

$$\begin{aligned} & N_1(\nu;\Theta_1) + N_2(\nu;\Theta_2) \\ &= N_1(\nu;\Theta_1) + N_2(\nu^\star;\Theta_2) \\ &= N_1(\nu;\Theta_1) - N_1(\nu^\star;\Theta_1) + V(\nu^\star) \\ &= V(\nu^\star) + \frac{1-\rho}{\eta}\int_{\nu^\star}^{\nu} \nu \frac{\partial^2 \tilde{g}}{\partial \nu^2}(\nu)\, d\nu \\ &= V(\nu) \end{aligned}$$

**(i).** Assume $\nu_1^\star = \nu_2^\star$. Then combining the two Equations of Lemma 6.15 shows that $\nu_1^\star$ is the solution of the Equation of Lemma 6.17. Thus $\nu_1^\star = \nu_2^\star = \nu^\star$. Similarly, it can be proven that $\tilde{g}_a(\nu_M) = \tilde{g}_b(\nu_M)$ and consequently that $\tilde{g}_a = \tilde{g}_b$ as solutions of the same differential equations with the same boundary conditions. Applying (ii) or (iii) leads to the result. □

Appendix C

# Proofs of Chapter 7

**Proposition 7.5.** *Assume given a state of the supply* $(\boldsymbol{\tau}_{RC}, \boldsymbol{p}_{RC})$ *and consider a user category* $c$ *with convex schedule delay cost function* $D$ *such that* $D(0) = 0$. *If two elements* $(r_1, h_1, h_p^1)$ *and* $(r_2, h_2, h_p^2)$ *of* $\mathcal{S} \times \mathcal{H}_p$ *are such that* $(r_i, h_i)$ *is the solution of the user optimization program for* $(c, h_p^i)$, $h_p^1 \leq h_p^2$ *and* $h_1 + \tau_{r_1 c}(h_1) \geq h_2 + \tau_{r_2 c}(h_2)$ *then:*

$$g(h_1, \tau_{r_1 c}(h_1), p_{r_1 c}(h_1) | c, h_p^i) = g(h_2, \tau_{r_2 c}(h_2), p_{r_2 c}(h_2) | c, h_p^i) \text{ for } i = 1 \text{ or } 2$$

*Proof of Proposition 7.5.* Since $(r_1, h_1, h_p^1)$ and $(r_2, h_2, h_p^2)$ are optimal:

$$\nu \tau_{r_1 u}(h_1) + p_{r_1 c}(h_1) + D(h_1 - h_p^1) \geq \nu \tau_{r_2 c}(h_2) + p_{r_2 c}(h_2) + D(h_2 - h_p^1) \quad (3.1)$$

and

$$\nu \tau_{r_2 c}(h_2) + p_{r_2 c}(h_2) + D(h_2 - h_p^2) \geq \nu \tau_{r_1 u}(h_1) + p_{r_1 c}(h_1) + D(h_1 - h_p^2) \quad (3.2)$$

Combining the two latter equations gives:

$$D(h_2 - h_p^2) + D(h_1 - h_p^1) \geq D(h_2 - h_p^1) + D(h_1 - h_p^2) \quad (3.3)$$

Then, since $D$ is convex and

$$\bar{h}_1 - h_p^1 < \bar{h}_1 - h_p^2 < \bar{h}_2 - h_p^2$$

it comes:

$$\frac{D(\bar{h}_1 - h_p^1) - D(\bar{h}_1 - h_p^2)}{h_p^2 - h_p^1} \leq \frac{D(\bar{h}_1 - h_p^1) - D(\bar{h}_2 - h_p^2)}{(\bar{h}_1 - h_p^1) - (\bar{h}_2 - h_p^2)}$$

Similarly as

$$\bar{h}_1 - h_p^1 < \bar{h}_2 - h_p^1 < \bar{h}_2 - h_p^2$$

it comes:

$$\frac{D(\bar{h}_1 - h_p^1) - D(\bar{h}_2 - h_p^1)}{\bar{h}^1 - \bar{h}^2} \leq \frac{D(\bar{h}_1 - h_p^1) - D(\bar{h}_2 - h_p^2)}{(\bar{h}_1 - h_p^1) - (\bar{h}_2 - h_p^2)}$$

Combining the two inequalities yields to:

$$D(h_1 - h_p^1) + D(h_p^2 - h_2) \leq D(h_2 - h_p^1) + D(h_1 - h_p^2) \tag{3.4}$$

And thus from (3.4) and (3.3):

$$D(h_1 - h_p^1) + D(h_2 - h_p^2) = D(h_2 - h_p^1) + D(h_1 - h_p^2)$$

Which yields to:

$$\begin{aligned} g \quad & (h_1, \tau_{r_1 u}(h_1), p_{r_1 c}(h_1)|c, h_p^1) + g(h_2, \tau_{r_2 c}(h_2), p_{r_2 c}(h_2)|c, h_p^2) \\ = \quad & g(h_1, \tau_{r_1 u}(h_1), p_{r_1 c}(h_1)|c, h_p^2) + g(h_2, \tau_{r_2 c}(h_2), p_{r_2 c}(h_2)|c, h_p^1) \end{aligned}$$

Using (3.1) and (3.2), we have the result. □

**Theorem 7.6** (On the order of arrival)**.** *Consider a DUE problem with atomless demand $\boldsymbol{X}_C^p$. Let $\boldsymbol{D}_C$ be a dynamic user equilibrium. Then there exists a dynamic user equilibrium $\boldsymbol{D}'_C$ such that the arrival distributions $(\bar{D}'_c)_{c\in C}$ are symmetric and that the symmetric reductions of $(\bar{D}'_c)_{c\in C}$ are non decreasing. Moreover for each category $c$ the marginal of $D'_c$ and $D_c$ on $\mathcal{S}$ are the equal.*

*Proof of Theorem 7.6.* Let $\boldsymbol{D}_C = (D_c)_{c\in C}$ be a dynamic user equilibrium and consider the associated arrival distributions $(\bar{D}_c)_{c\in C}$ as well as $(\bar{X}_{rc})_{r\in R, c\in C}$, the marginals of $(\bar{D}_c)_{c\in C}$ on $\mathcal{H} \times \{r\}$. The quantity $\bar{X}_{rc}$ represents the cumulated flow of the users of category $c$ at the exit of route $r$.

For any $c$, let $\bar{H}_c := \bar{X}_{rc}^{-1} \circ X_c^p$. Here $\bar{X}_{rc}$ and $X_c$ are seen as absolutely continuous functions rather than measures;. Note that $\bar{H}_c$ is well defined as $X_c^p$ is atomless. Then define $\bar{D'}_c$ as :

$$\bar{D'}_c(\{r\} \times I \times J) := \bar{X}_{rc}(I \cap \bar{H}_c(J)) \tag{3.5}$$
$$\text{for all } r \text{ in } R,\ I \subset \mathcal{H} \text{ and } J \subset \mathcal{H}_p$$

It yields:

$$\begin{aligned} \bar{D'}_c(R \times \text{graph } \bar{H}_c) &= \sum_r \bar{X}_{rc}(\mathcal{H}_p) \quad \text{(by construction)} \\ &= X_c^p(\mathcal{H}_c) \end{aligned}$$

Thus $\bar{D'}_c$ is symmetric. It is also straightforward that the marginals of $D'_c$ and $D_c$ on $\mathcal{S} = \mathcal{H} \times R$ are the same.

**Let us show that $(D'_c)_{c\in C}$ is a dynamic user equilibrium.**

Denote $(\boldsymbol{\tau}_{RC}, \boldsymbol{p}_{RC}) = F_S(\boldsymbol{X}_{RC})$ the travel times arising from the distributions $(D_c)_{c\in C}$ and $(D'_c)_{c\in C}$. By definition of an equilibrium, there exists a subset $E$ of $\mathcal{S} \times \mathcal{H}_p$ such that:

- $D_c(E) = X_c^p(\mathcal{H}_c)$
- For all $(r, h, h_p) \in E$, we have
$$g(h, \tau_{rc}(h), p_{rc}(h)|c, h_p) = \min_{(h,r)} g(h, \tau_{rc}(h), p_{rc}(h)|c, h_p)$$

Consider the set $\text{proj}_{\mathcal{S}}\ E$. By definition $D'_c(\text{proj}_{\mathcal{S}}\ E \times \mathcal{H}_p) = G(\mathcal{H}_c)$. Therefore there exists a closed subset $F$ of $\text{proj}_{\mathcal{S}}\ E \times \mathcal{H}_p$ which is the smallest (in the sense of inclusion), such that $D'_c(F) = X_c^p(\mathcal{H}_c)$ (see Hildenbrand, 1974, pp 49). Note that $\text{proj}_{\mathcal{S}}\ E = \text{proj}_{\mathcal{S}}\ F$.

Let $(r_1, h_1, h_p^1)$ an element of $F$. Then there exists $r_2, h_2$ and $h_p^2$ such that:

- $(r_2, h_2, h_p^1) \in E$ as $(r_1, h_1, h_p^1) \in F$ and $\text{proj}_{\mathcal{S}}\ E = \text{proj}_{\mathcal{S}}\ F$;
- $(r_2, h_2, h_p^2) \in F$ as $(r_2, h_2, h_p^1)$ in $E$ and $\text{proj}_{\mathcal{S}}\ E = \text{proj}_{\mathcal{S}}\ F$.

Assume without loss of generality that $h_p^1 \geq h_p^2$. Then, as $\bar{H}_c$ is an increasing function, $\bar{H}_c(h_p^1) = h^1 + \tau_{r_1c}(h^1) \leq \bar{H}_c(h_p^2) = h^2 + \tau_{r_2c}(h^2)$. Using proposition 7.5, it comes that:

$$\begin{aligned} g(h_1, \tau_{r_1c}(h_1), p_{r_1c}(h_1)|c, h_p^1) &= g(h_2, \tau_{r_2c}(h_2), p_{r_2c}(h_2)|c, h_p^1) \\ &= \min_{r,h} g(h, \tau_{rc}(h), p_{rc}(h)|c, h_p^1) \end{aligned}$$

It has been shown that $F$ is such that:

- $D'_c(F) = X^p_c(\mathcal{H}_c)$
- For all $(r, h, h_p) \in F$, we have
$$g(h, \tau_{rc}(h), p_{rc}(h)|c, h_p) = \min_{(h,r)} g(h, \tau_{rc}(h), p_{rc}(h)|c, h_p)$$

Hence the result. □

Appendix D

# Loading traffic on a network of bottlenecks: an event based approach

Dynamic user equilibrium computation aims to find in a network subject to congestion, time-varying traffic flows on routes that are consistent with the route travel costs. Thus, computing the travel times from given route volumes is both essential, as it's the main step in estimating the travel costs, and challenging, as the problem has both a temporal and network dimension. The problem boils down to derive the traffic volumes on each arc from the route volume vector. It is usually referred to as the *Dynamic Network Loading Problem* (DNLP).

Although Friesz et al. (1993) pointed out its importance for the analytical formulation of dynamic assignment models, the literature is quite restricted. Most of the existing solutions rely on simulations. They can be either microscopic (*e.g.* DYNASMART in Mahmassani et al., 1995), with an explicit representation of users behaviours on arc and nodes, or macroscopic, the demand being divided into packets of users . Analytical forms of the DNLP are not as frequent. Wu et al. (1998) first formulated the loading problem as a system of functional equations and derived a solution method based on a finite dimensional approximation. Xu et al. (1999) and later Rubio-Ardanaz et al. (2003) proposed a radically different approach that can be considered as an event based simulation, and showed it improves significantly the computation speed. Both methods apply to volume-delay travel time models. Yet volume-delay travel times have been shown to be generally unphysical (Daganzo, 1995) and an important number of operational assignment models rather assume bottleneck travel times (*e.g.* Kuwahara and

Akamatsu, 1993; Leurent, 2003*a*). DNLP for the bottleneck model has never been treated explicitly, despite the fact it is regarded as an important category of models combining analytical simplicity, computational robustness, and experimental correctness.

In this chapter, we propose a general solution paradigm, inspired by discrete event simulation, and applied it to a network of bottlenecks. The first section presents the global philosophy of the solution method, while the third gives a formal statement of the algorithm. Finally the fourth section gives an example of applications on a network of bottlenecks and numerical experiments.

In Chapter 3, we proposed an original formulation for the DNLP and shown that under five assumptions on the arc travel time models existence and uniqueness of the problem was guaranteed. As the proof is constructive, a computation algorithm can be derived

# 1 Philosophy of the solution method

## 1.1 Statement

The notations of this chapter are essentially the one of Chapter 3. However, instead of seeing the route cumulated flows as measures on an interval $I$, we will rather consider increasing continuous function $X$ with the following meaning : $X(h)$ is the quantity of traffic that have passed through a point since the beginning of $I$. In other words, $X([\min I, h])$ is now denoted $X(h)$. Similarly arc cumulated flows are seen as increasing continuous function $Y$ on $\mathbb{R}$. A vector of cumulated flows $\boldsymbol{X}^R = (X^r)_{r\in R}$ is called a route volume vector and a vector of cumulated flows $\boldsymbol{Y}_A = (Y_a)_{a\in A}$ is called an arc volume vector.

Apart from this slight notational changes, the dynamic traffic loading problem is stated as in Chapter 3.

**Definition D.1** (Dynamic traffic loading problem)**.** *For given route cumulated flows at origin $\boldsymbol{X}^R = (X^r)_{r\in R}$, find arc cumulated volumes $\boldsymbol{Y}_A = (Y_a)_{a\in A}$ such that there exists a collection of route vector flow $(\boldsymbol{Y}_a^R)_{a\in A} = (Y_a^r)_{a\in A, r\in R}$ satisfying the system:*

$$Y_a = \sum_{r\in R:\, a\in r} Y_a^r \tag{4.1}$$

*and for all* $r = a_1, \ldots, a_n \in R$

$$\begin{array}{rcll} Y^r_{a_1} & = & X^r & \text{(i)} \\ Y^r_{a_i} \circ H_{a_{i-1}}[Y_{a_{i-1}}] & = & Y^r_{a_{i-1}} \quad \text{for } i = 2, \ldots, n & \text{(ii)} \\ Y^r_a & = & 0 \quad \text{if } a \notin r. & \text{(iii)} \end{array} \tag{4.2}$$

It was shown that the Dynamic traffic loading problem admits a unique condition under 5 assumptions on the arc travel time models (Assumptions I-V), namely continuity, no infinite speed, finiteness, strict fifoness, causality.

## 1.2 Restrictive assumptions

The proof in the previous chapter is constructive and thus gives an algorithm to load traffic on a road network with arc travel time models satisfying to the Assumptions (I-V). In a few words the general idea is to proceed recursively. Assume known a collection of cumulated flows $(Y^r_a)_{a\in A, r\in R}$ satisfying equations (1-3) until the instant $h$, then by Assumption III, IV and V, you can deduce a new collection cumulated flows $(Y^r_a)_{a\in A, r\in R}$ satisfying the same equations until $h + t_{min}$. Assumption III guarantees the termination of the recursion. In the general case this algorithm is possibly the most efficient way of solving the problem and in fact in the literature most of the existing algorithms follow more or less this same pattern. Yet by slightly restricting the problem, it can be importantly simplified thus leading to a more efficient solution method.

In this paper only route cumulated volumes at origins which are continuous piecewise linear functions of time are considered. In addition, arc travel time models are assumed to lead to piecewise linear travel time functions when applied to continuous piecewise linear route volume vectors.

The primitives manipulated under these assumptions are piecewise linear (PWL) functions. Note that they can easily be encoded under the form of an ordered list of elements $X = (h_i, X_i, x_i)_i$ where $X_i$ is the image of $h_i$ by the PWL function $X$ in and $x_i$ is its derivative on the right. Each element of the list is called a piece. Note that under this formalism it is easy to define the operations of linear combination, composition and inverse. With an adequate implementation there are essentially equivalent to a list traverse and thus in $O(n)$ where $n$ is the number of pieces of the function considered.

## 1.3 Consequences for the loading problem

With the PWL assumptions, the main quantities of the problem are piecewise linear functions of the time. Let us focus on the cumulated flows $(Y_a^r)_{a\in A,r\in R}$ and consider the set of instants $h_i$ such that there exists a triplet $(h_i, y_i, s_i)$ that belongs to a cumulated flows $Y_a^r$. They will be referred to as the *critical instants*. Now assume the cumulated flows $(Y_a^r)_{a\in A,r\in R}$ are known until $h$, *i.e.* that the $(Y_a^r|_h)_{a\in A,r\in R}$ are known. In terms of PWL format it means that all elements $(h_i, X_i, x_i)$ of $Y_a^r$ such that $h_i \leq h$ are known. Then if one could find the first critical instant $h'$ after $h$, as well as the concerned cumulated volume and its new slope, $(Y_a^r|'_h)$ can be deduced and the process can be iterated until completion.

Informally that is just saying that instead of seeing the cumulated flows as functions of the times, we see them as a sequence of transitions from one slope to another occurring at critical instants. The algorithm we proposed is simply to go from critical instants to critical instants and correctly update the values of $(Y_a^r)_{a\in A,r\in R}$. Our proposition is to formalize this general idea under a discrete event system. Each event corresponds to a change of slope and thus occurs at a critical time.

## 1.4 Example on a simple case

We are going to consider a simple example using the bottleneck travel time model with constant capacity. Assume a simple network of bottlenecks with two origins, $O_1$ and $O_2$, two destinations $D_1$ and $D_2$ and five arcs denoted $a_i, i = 1..5$. Only two routes are available, $r_1 = a_1, a_3, a_4$ connects $O_1$ to $D_1$ while $r_2 = a_2, a_3, a_5$ connects $O_2$ to $D_2$. The only bottleneck is on $a_3$ with a capacity of $k = 1000 uvp/hour$. The free flow travel times of arc $a_1$, $a_2$ and $a_3$ are given respectively by $t_{0,a_1} = 1$ hour, $t_{0,a_2} = 1.5$ hour and $t_{0,a_3} = 0$. The network and its characteristic values are depicted in Figure D.1.

Consider a route volume vector $\boldsymbol{X} = (X^{r_1}, X^{r_2})$ defined on $I = [6:00, 14:00]$. $X$ is given by its derivative $x$. During $[6:00, 9:00)$, the flow entering $r_1$ is $x^{r_1} = 1500 uvp/hour$ while the flow on route $r_2$ is $x^{r_2} = 500$ uvp/hour. On $(9:00, 14:00]$ we have $x^{r_1} = x^{r_2} = 250$ uvp/hour. The resolution can be made iteratively.

1. At 6 : 00 all the flows of the network are zero except on arc $a_1$ and $a_2$ where $y_{a_1} = x^{r_1} = 1500$ uvp/hour and $y_{a_2} = x^{r_2} = 500$ uvp/hour. This

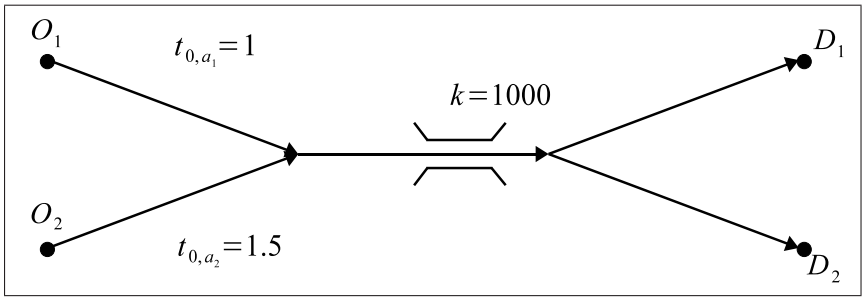



Figure D.1: A simple example of traffic loading: the network

stay so until the route flows reaches the end of arc $a_1$ and $a_2$. This happens at $h_1 = 6:00 + t_{0,1} = 7:00$ for arc $a_1$ and at $h_2 = h_A + t_{0,2} = 7:30$ for $a_2$.

2. At $h_1$, the incoming flow on $a_3$ changes from $y_{a_3} = 0$ to $y_{a_3} = 1500$ uvp/hour. As the outgoing flow from $a_3$ is bounded by the bottleneck in capacity, there is an exit flow of $y^-_{a_3} = k = 1000$ uvp/hour. A queue began to grow at the rate of $\frac{dQ_{a_3}[Y_{a_3}]}{dh} = 1500 - 1000 = 500$ uvp/hour. As all users on $a_3$ are following route $r_1$ for the moment, the exit flow is entirely disgorged on arc $a_4$ and $y_{a_4} = 1000$ uvp/hour.

3. At $h_2$, the flow entering $a_3$ switches to $y_{a_3} = 2000$ uvp/hour. The queue growing rate is now $\frac{dQ_{a_3}[Y_{a_3}]}{dh} = 2000 - 1000 = 1000$ uvp/hour and the travel time on that arc is $t_{a_3}[Y_{a_3}](h_2) = (Q_(a_3)[Y_{a_3}](h_2))/k = 0.25$ hour. Consequently the first user following route $r_2$ to exit arc $a_3$ will arrive on $a_5$ at $h_3 = 7:45$.

4. At $h_3$, the new incoming flows of arc $a_4$ and $a_5$ are $y_{a_4} = 750$ uvp/hour and $y_{a_5} = 250$ uvp/hour respectively. The change in route flow at $9:00$ provokes changes in the incoming flow of arc $a_3$ at instants $h_4 = 9:00$ and $h_5 = 9:30$.

5. At $h_4$, $y_{a_3} = 1250$ uvp/hour and $\frac{dQ_{a_3}[Y_{a_3}]}{dh} = 250$. The travel time is now $t_{a_3}[Y_{a_3}](h_4) = 1.5$ hours. The next change in the exit flow will be at $h_6 = 10:30$.

6. At $h_5$, $y_{a_3} = 500$ uvp/hour and $\frac{dQ_{a_3}[Y_{a_3}]}{dh} = -500$. The travel time is now $t_{a_3}[Y_{a_3}](h_5) = 1.625$ hours. The exit flow will change at $h_7 = 11:07$. The queue is now decreasing and will be empty by $h_8 = 12:45$.

7. At $h_6$, the entrance flows on $a_4$ and $a_5$ are $y_{a_4} = 1000/3$ and $y_{a_5} = 2000/3$ respectively.

8. At $h_7$, the entrance flows on $a_4$ and $a_5$ are $y_{a_4} = y_{a_5} = 500$ uvp/hour.

9. At $h_8$, the queue is completely cleared and the entry flows on arc $a_4$ and arc $a_5$ corresponds to the exit flow of arc $a_1$ and $a_2$ *i.e.* $y_{a_4} = y_{a_5} = 250$ uvp/hour.

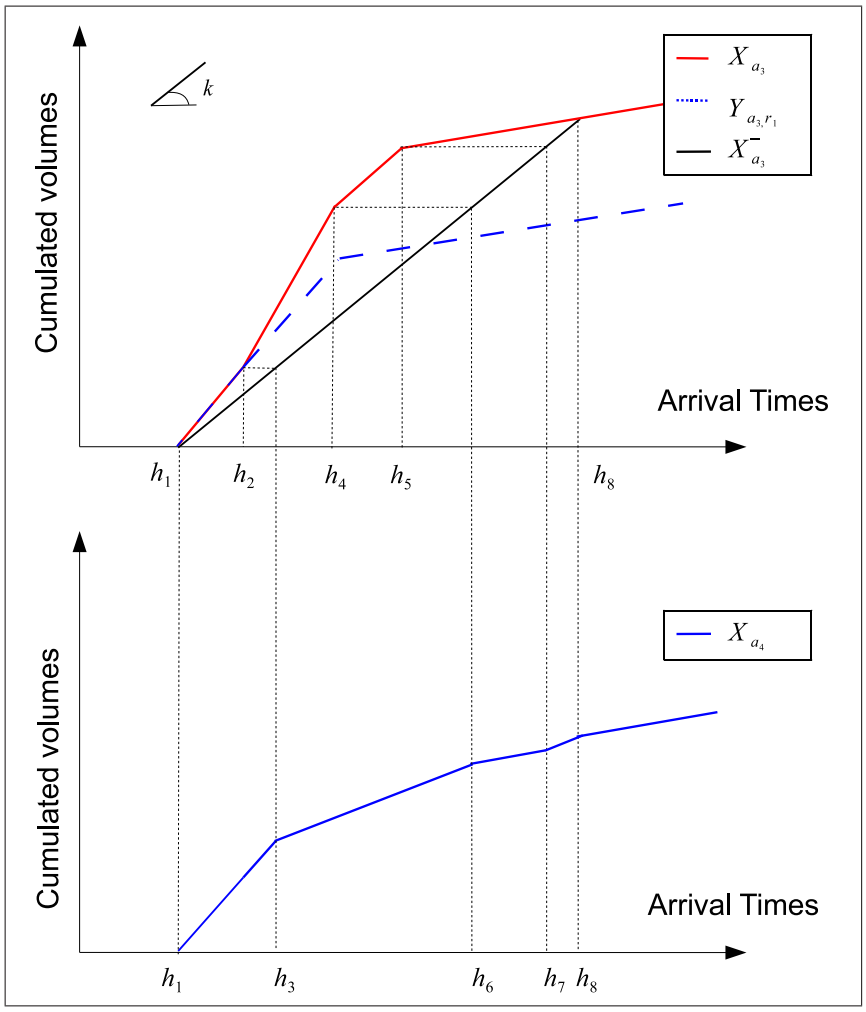



Figure D.2: A simple example of traffic loading: the loaded flows

Figure D.2 depicts the solutions of the loading problem by representing the cumulated volume $Y_{a_3}$ at the entrance of arc $a_3$, the cumulated flow $Y^{r_1}_{a_3}$ at the entrance of arc $a_3$ following route $r_1$, the cumulated volume $Y^-_{a_3}$ at the exit

of $a_3$ (Figure D.2, top) and the cumulated flow $Y_{a_4}$ at the entrance of arc $a_4$ (Figure D.2, bottom). This example, albeit simple, is quite instructive. First, relatively simple inputs in terms of both networks and route flows can lead to much more complicated arc flows through the interaction at bottlenecks. Second it is reasonably easy to deduce the impacts of a change in an arc incoming flow (*i.e.* an event with our terminology) on the change of flows on the downstream arcs. Informally the mechanism is the following: from each event can be deduced to a sequence of other forthcoming events that needs to be treated chronologically. The main difficulty is to correctly coordinate the actualization of the arc entering flows *i.e.* to handle each event in the right (chronological) order. The algorithm presented in the following essentially addresses this issue.

# 2 General loading algorithm statement

## 2.1 Data structures

First let us define the proper data structure to model the problem under a discrete event system. Basically, we need to be able to describe a state of the system, to formalize the concept of events and to treat events.

**Instantaneous flow vectors and events.** A flow vector is a vector of instantaneous flows $\boldsymbol{y}_R = (y^r)_{r\in R}$ and has the following physical interpretation: it represents the superposition of the flow of users following different route in a given point at a given instant. The sequence of flow vectors $S^{\text{ext}} = (y_a)_{a\in A}$ is called the external state of the system, and represents flows, decomposed according to the followed route, at the entrance of each arc of the network. The term external refers to the fact that this description only focuses on arc incoming flows and totally ignores what's happening inside the links, which we will later refer to as the internal state of a system. Despite that, its knowledge over the simulation period (*i.e.* for every instants) is exactly what's required to solve the DNLP. Extending $S^{\text{ext}} = (y_a)_{a\in A}$ description over the whole period of simulation is hence of interest and so we introduce the concept of event as a change of flow vector at a specific instant and on a specific node.

Formally, an event is defined as:

**Definition D.2** (Event)**.** *An event is a pair* $(h, \boldsymbol{e})$ *where:*

- *h is a clock time*

- $\boldsymbol{e}$ *is a map on* $A \times R$ *such that* $\boldsymbol{e}(a, r)$ *is either an instantaneous flow or the empty set* $\emptyset$.

**Arc event functions.** Let's now focus on the arc description. At this point of our exposition, we remain general and only expose how to represent travel time model in an event based perspective. First, one needs to be able to describe the state of an arc. Physically the state of an arc describes at an instant $h$ the traffic flows over the whole arc. In other terms, it is what you cannot see by solely looking at the arc incoming and outgoing flows. In a computational perspective, we would like the knowledge of the state of an arc to be enough to compute all of the future values of the arc travel time and outgoing flows over an infinite time horizon, assuming the route flow vector y at entrance remains constant. According to the causality assumption, in the general case ? $Y_a|_h$ is enough to compute the travel time at $h$. Consider the route flow vector $Y'_a$ obtained by considering for each route ? $Y^r_a|_h$ and prolonging it using the instantaneous flow following the corresponding route $x^r$. The knowledge of all the route volumes thus obtained is then sufficient to compute the travel time for any instants assuming that the incoming flow remains constant. This discussion leads to choose to represent the state of an arc simply by a route volume vector $\boldsymbol{Y}^R$. The inner state of the system is then naturally defined as a collection of route volume vectors $S^{in} = (Y_a)_{(a \in A)}$, one for each arc of the network.

We denote the operation of prolonging a cumulated volume $X$ by an instantaneous flow $X \oplus_h x$. Note that in PWL format, denoting $X = (h_i, X_i, x_i)_{i=1...n}$, it is equivalent to take the sequence of pieces $(h_i, X_i, x_i)$ such that $h_i < h$and to add the piece $(h, X(h), x)$.

**Definition D.3** (Event functions)**.** *For each arc travel time model* $t_a$*, the following functions are defined:*

- ***The next event function*** $F_a : (Y, h) \to (h', \boldsymbol{e})$ *, where* $\boldsymbol{e}$ *is the event representing the first change in slope in the outgoing route flows* $Y^r_{a,-} := Y^r_a \circ H_a[Y]$ *after h. Denote* $h'$ *the instant when this change occurs. Then for all* $r : a \in r$*, considera*$'$ *the first arc after a in r and define* $\boldsymbol{e}(a', r) := y^r_{a,-}$*. If* $h'$ *does not exist, then let* $h' := +\infty$ *and* $\boldsymbol{e} := \emptyset$.

- ***The handling function*** $U_a : (\boldsymbol{Y}_1, \boldsymbol{y}, (h, \boldsymbol{e})) \to \boldsymbol{Y}_2$ *such that* $Y_r^2 := Y_r^1 \oplus_h \boldsymbol{e}(a, r)$ *if* $\boldsymbol{e}(a, r) \neq \emptyset$ *and* $Y_2^r := Y_1^r$ *otherwise.*

Note that the expression $(+\infty, \emptyset)$ encodes the null event that indicates that no event is generated by a function $F_a$.

For specific travel time models, such as bottleneck ones, using a route vector flow to describe the sate of an arc might not be the most suitable choice. In this situation the event functions will need to be adapted to feet this new model. Yet the global framework of the algorithm will remain the same. This will be discussed later in the application to a network of bottlenecks.

The precise statement of our event-based loading algorithm is given below. The inputs are simply the arcs described by their event functions and the route volumes at origin described by a collection of events. The outputs are route volume vectors, one for each arc, representing the cumulated volumes at the entrance. The algorithm can be summarized as such. Consider the list of event formed by the merge of the events at origins and the events on arcs and remove the first event. Then use the handling function to update the state of the arc concerned with the event. Finally compute the new arc event list using the next event functions of each arc of the network.

**Algorithm D.1** `loadingTraffic`$((H_a),(F_a),E^O)$

**Inputs:** - A list of arc $A = (a_1, \ldots, a_n)$ together with the corresponding functions $H_a$ and $F_a$ for each $a \in A$

**Inputs:** - A list of events $E^O = (h_1, \boldsymbol{e}_1), \ldots, (h_i, \boldsymbol{e}_i), \ldots$

**Outputs:** - A collection of route flow vector $(\boldsymbol{Y}_a)_{a \in A}$

**Initialize** $y_a$ and $Y_a$ to 0 for all $a \in A$ ,$E^A$ to the empty list and $h$ to a suitable initial instant.

**While** $E^O \cap E^A \neq \emptyset$

Get next event from $E^O \cap E^A$ and **Set** it to $(h, \boldsymbol{e})$.

**For all** $(a, r) : (\boldsymbol{e}(a, r) \neq \emptyset)$, set $y_a^r := \boldsymbol{e}(a, r)$

**Foreach** arc $a : \exists r : (\boldsymbol{e}(a, r) \neq \emptyset)$,

**Set** $Y_a := U_a(Y_a, (h, \boldsymbol{e}))$

**If** $F_a(Y_a, (h, \boldsymbol{e})) \neq (+\infty, \emptyset)$, add it to $E^A$

**End For**

**End While**

# 3 Application to a network of bottlenecks

## 3.1 General presentation

In this section we consider a network of bottlenecks where each arc a is described by two parameters: its free flow travel time $t_{0,a}$ and its exit capacity $k_a$. We are going to apply to this network our general algorithm. To do so divide each arc in two parts: the free flow part and the bottleneck part. It is this new network we are going to consider and thus two types of arc time models and event functions have to be defined.

The treatment of the free flow part is straightforward and can be achieved by directly applying the general method presented above. Concerning the bottleneck part, a modification to the next event function is presented below and allows accelerating computation by storing slightly more information in the bottleneck state.

**Bottleneck part.** As precised earlier, in this case using solely a route flow vector $\boldsymbol{X}$ to describe the state of a bottleneck is not the best choice from a computational perspective. A more suitable choice is to have some information about the queue evolution. To do so let us add to the state of a bottleneck the function $Q$. The quantity $Q$ is a positive function of the time representing the queue volume with respect to the time.

It is now necessary to adapt the event functions in order to exploit the additional information given by $Q$. Given an bottleneck state $(Y, Q)$, how can one compute the next event? Assume the queue is not empty at an instant $h$. From the analytics exposed above, two cases can arise:

1. The outgoing flow change because of a previous change in the incoming flow. Denoting the corresponding incoming route flows $y^r$, the new outgoing flows are $y^r_- := y^r/(\sum_{r' \in R} y^{r'})k_a$. Finding this instant corresponding to the change in the incoming flow boils down to find an instant $h'$ such that $h' + Q(h)/k_a < h$and the derivative of a route volume $Y^r$ changes in $h'$.

2. The outgoing flows change because the queue vanishes. The new outgoing flows are simply the incoming flows $y^r_- := y^r$. Finding the instant h' occurs is straightforward knowing $Q$.

The event functions for bottleneck are precisely defined below.

**Definition D.4** (Event functions for bottlenecks)**.** *Denote* $\boldsymbol{y}_a = (y_(a,r))_{r\in R}$ *the incoming route flow on arc* $a$ *and* $n(a,r)$ *the first arc after* $a$ *in route* $r$*. The events functions of a bottleneck are then defined as follow.*

- ***The next event function*** $F_a : (\boldsymbol{Y}, Q, h) \to (h', \boldsymbol{e})$ .

  **Case 1:** $Q(h) = 0$

  - *If* $y^r = dY^r/dh$ *then* $h' := +\infty$ *and* $\boldsymbol{e} := \emptyset$.
  - *else* $h' := h$ *and for all* $r$ *define* $\boldsymbol{e}(n(a,r),r) := y^r$.

  **Case 2:** $Q(h) \neq 0$

  - *Let* $h'$ *the last change in slope of* $Y^r$ *of such that* $h' + Q(h)/k_a < h$*. Then for any* $r \in R$*, let* $y_r$ *be the right derivative of* $Y^r$ *in* $h'$ *and define* $\boldsymbol{e}(n(a,r),r) := \frac{y^r}{\sum_{r'\in R} y^{r'}} k_a$.
  - *If there is no such* $h'$*, let* $h'$ *be the first instant such that* $Q(h') = 0$*. Then for any* $r \in R$*, let* $y^r$ *be the right derivative of* $Y^r$ *in* $h'$ *and define* $\boldsymbol{e}(n(a,r),r) := y^r$.

- ***The handling function*** $H_a : (Y_1, Q_1, (h, \boldsymbol{e})) \to (Y_2, Q_2)$ *such that* $Y_2^r := Y_1^r \oplus_h \boldsymbol{e}(a,r)$ *and* $Q^2 = Q^1 \oplus_h \sum_{r\in R} \frac{dY_2^r}{dh} - k_a$

We claimed that this latter implementation of the event function is more efficient than the former one. Why is that? The general implementation proposed to seek the next event by computing for each call of the function next event the travel time for the current state of the arc. Consequently it does not use at all the information gathered through the previous calls of this function. Yet for a bottleneck model this requires the integration of a first order differential equation of PWL functions and thus a full list traverse. On the contrary, the adaptation for the bottleneck model exploits this information by keeping in memory the queue. The most computational intensive operation is to perform the operation described by the first item of case 2 in Definition D.4. Although in the worst case this operation also requires a list traverse, it tends to be a simple scan forward on the last pieces of $Q$.

## 3.2 Numerical illustration

In this subsection, a small but instructive example is presented. We consider the network of four arcs and two OD pairs presented in Figure D.3. Only the arcs $O_1 - D_2$ (arc 1) and $O_2 - D_1$ (arc 2) are subject to bottleneck congestion and they both have a capacity of k=2000 pcu/h. All the arcs have a free flow travel time of $t_0 = 1$. Only two routes are considered, the first one being $O_1 - D_2 - O_2 - D_1$ (route $r_1$) and the second one $O_2 - D_1 - O_1 - D_2$ (route $r_2$). The simulation period is $I = [5:00, 20:00]$ and the instantaneous flow on $r_1$ and $r_2$ are respectively a discretized gaussian shaped curve centered in 10 : 00 and a simple constant flow of 1000 pcu/h. The inputs are plotted in Figure D.4.

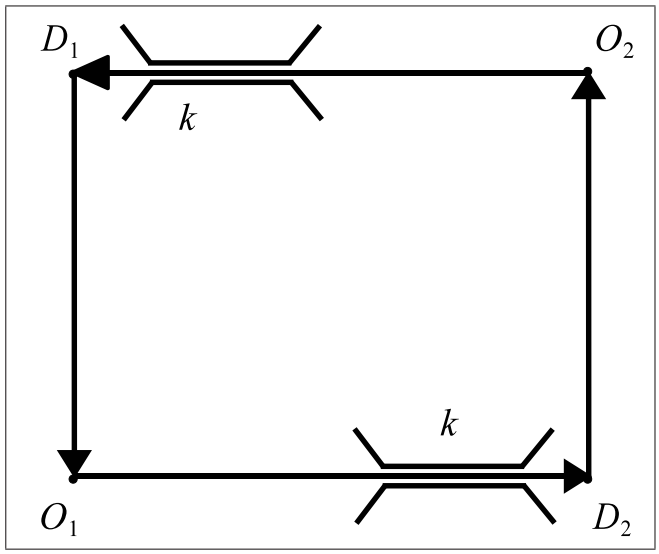



Figure D.3: Numerical illustration of traffic loading: the network

This network configuration is especially interesting and complex from an event based perspective. Assume that the instantaneous flow on route $r_1$ increases then the travel time on $a_1$ is going to increase and eventually the proportion of flows exiting $a_1$ and following route $r_1$ will also. This in turn results in a rise in $y_{a_3}^{r_1}$ and finally in a decrease in $y_{a_4}^{r_2}$ and in the incoming flow on $a_1$. The network thus acts as a sort of feedback loop, inducing a decrease in the flow on a route when the flow on the other route grows.

The travel times resulting from the loading of the traffic are plotted in Figure 3. The feedback effect exposed a few lines above can be seen on the travel time on $a_1$. It results in oscillations around the maxima of in travel

time. Also note the shape of the travel time on $a_3$ where two distinct maxima appear.

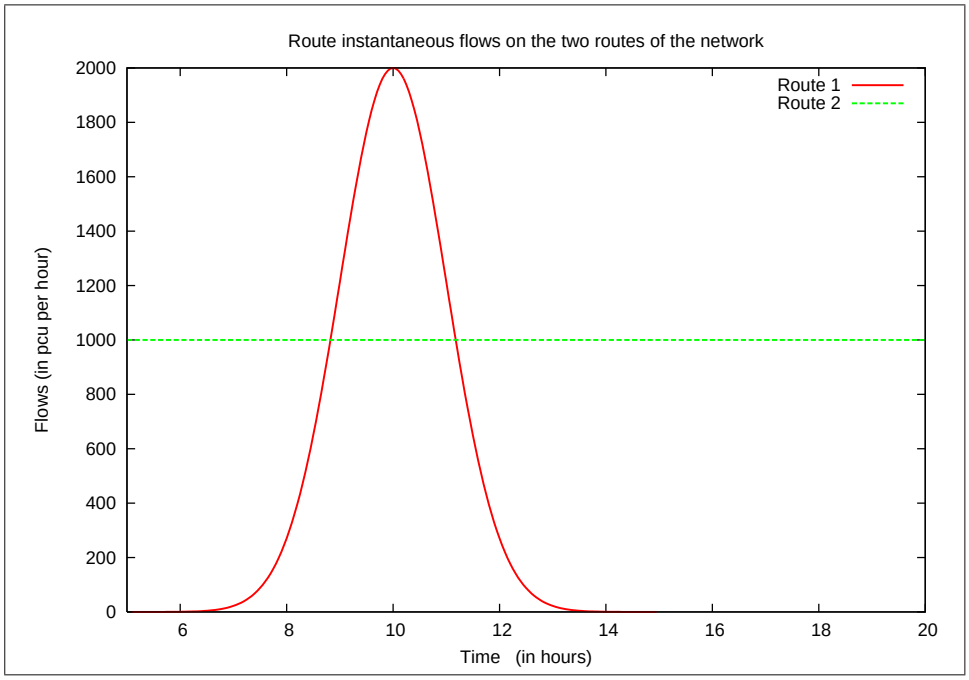



Figure D.4: A simple example of traffic loading: the inputs

## 3.3 Benchmark

The running time of the event based algorithm applied on a network of bottlenecks is clearly proportional to the number of events treated. Yet this latter quantity is impossible to compute *a priori*. In this subsection a small numerical experiment is conducted in order to get some insights about the sensitivity of the number of events with respect to the volumes at origin.

The setting is the following. A randomly generated network of 80 nodes and 200 arcs is considered. Then routes of 10 arcs are randomly generated, each of them assigned with a Gaussian shaped flows discretized in 20 pieces. This experiment has been conducted several times with a number of routes varying between 5 and 80.

Why is this numerical experiment relevant? One of the main applications of the DNLP is its integration in dynamic traffic assignment algorithms. Yet in most numerical schemes for dynamic traffic assignment, on progressively

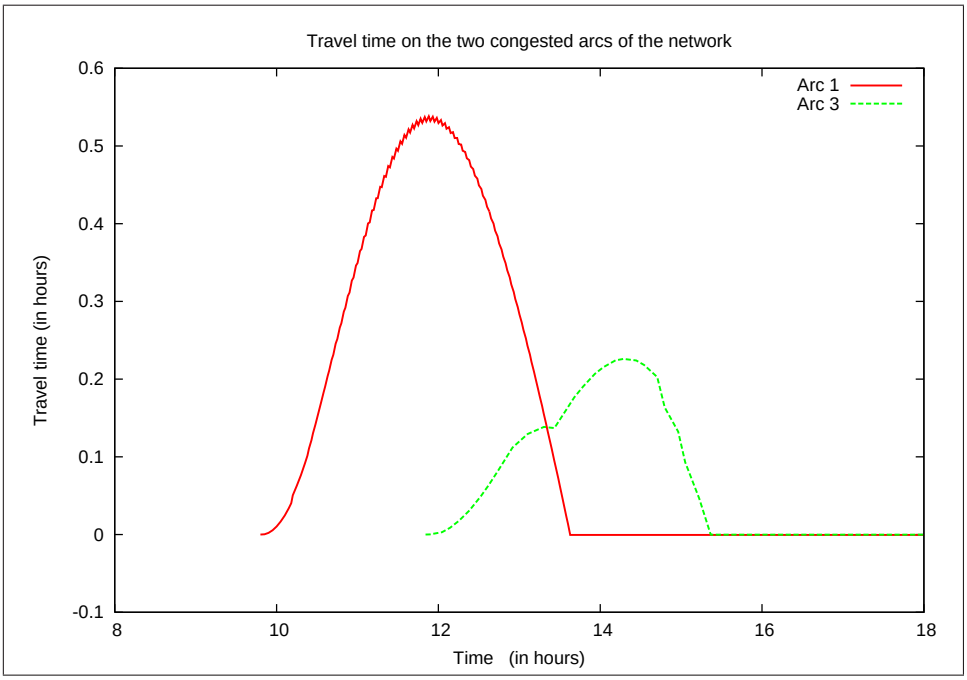



Figure D.5: A simple example of traffic loading: the results

discover new routes to serve an origin destination and assign part of the traffic on them. Thus the further the algorithm goes, the more routes are loaded with traffic. Another alternative would have been to try networks of different size. But in the DNLP, the size of the network, on the contrary to many other graph-based algorithms has little influence. The dimensioning quantity is rather the number of routes and the way they overlap themselves.

Figure D.6 shows the evolution of the number of events with the number of routes. At first the evolution is roughly linear. Around 50 routes the slopes quickly switch to a much higher value. Around 70 routes it seems that evolution becomes linear again.

A possible explanation for this behavior is given by the way the events propagate themselves over the network. When the number of routes assigned with traffic is low on a network, those routes tends not to intersect. Consequently the number of event is roughly the number of pieces of the volumes at origin times the average number of arcs on a route. However as the number of routes increases, more routes intersect and quickly an event occurring at given place of the network tends to propagate all over the network. This phenomenon is depicted on Figure D.7.

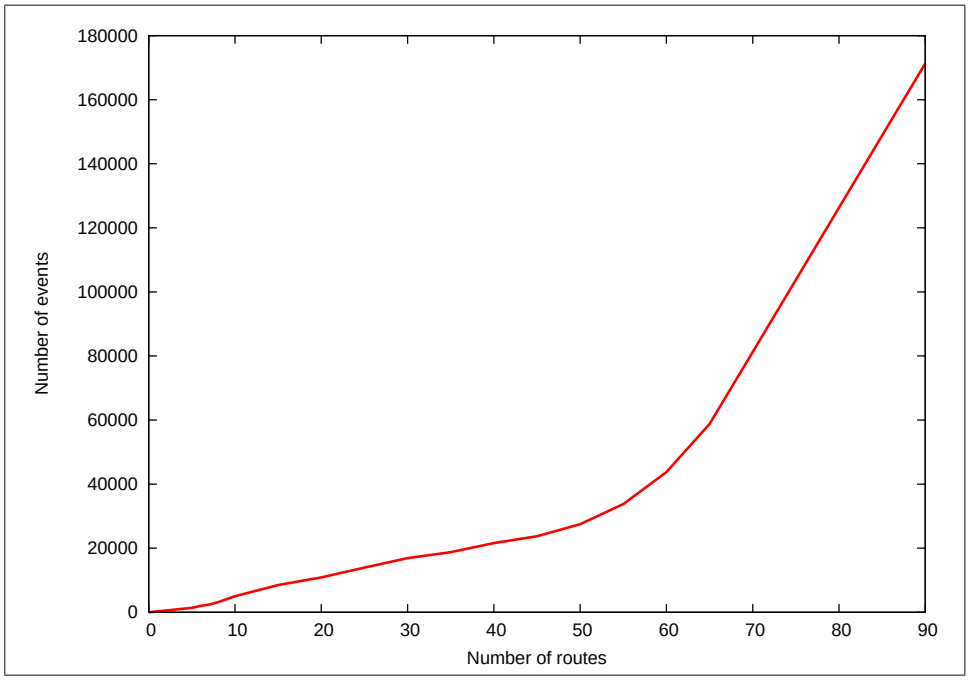



Figure D.6: Number of events as a function of the number of routes

Note that the number of events is a good proxy for the running time of our algorithm, but also of the actual "complexity" of the result of the dynamic loading. Indeed the number of event treated is essentially the number of pieces of the resulting cumulated flows on the arcs. In that perspective our results can be interpreted in two ways. On one hand, it is good news, as the running time seems to be asymptotically linear with the size of the route flow vector in input. But on the other hand, the practical number of events to deal with a reasonably small example is quite high. For real size networks with an important number of an origin-destination pairs, an efficient exact algorithm seems to be difficult. This is a strong argument for approximate loading procedure.

# Conclusion

In this chapter, a generic algorithm for the dynamic network loading problem has been presented and an application to a network of bottleneck has been presented. It was also an opportunity to gives a theoretical insight of the traffic flowing on a network in a dynamic context. Among our findings, we

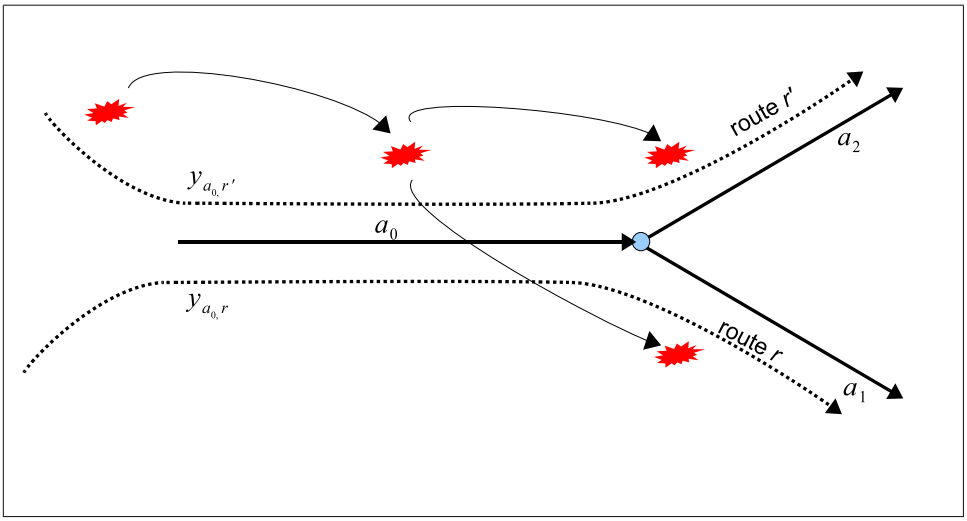



Figure D.7: Event Propagation in a DNLP. In this simple network, two routes are going through $a_0$. If an event occurs on route r', upstream arc $a_0$, it will cause an event on arc $a_0$ and thus on all downstream arcs of route r

have seen that the outputs of the loading problem quickly grow in complexity due to the complex interactions of the route flows on the networks.

Although the examples of application presented here are rather simple, this event based algorithm is an interesting first step toward a much more generic framework for dynamic network loading. The concept of events was here restricted to a change in the incoming flow of an arc. But one could introduce a wide variety of events modelling various physical phenomena. For instance queue spillback could be considered by adding an event type “arc a is full” and updating the upstream arcs in consequence. In the same order of idea dynamic traffic regulation schemes such as dynamic traffic regulation techniques could be modelled in an event based perspective. This offers vast possibilities for future works.

Appendix E

# LabDTA: a prototyping environment for dynamic user equilibrium computation

## 1 Main features

LabDTA stands for **L**aboratory for **D**ynamic **T**raffic **A**ssignment. It is a small toolbox that allows to test dynamic traffic assignment algorithms as well as algorithms to compute dynamic user equilibriums. The modelling framework proposed by LabDTA is essentially the one of LADTA, a dynamic traffic assignment model introduced by Leurent (2003*b*).

LabDTA is rather well-suited for rapid prototyping but is not intended to work with large scale networks. For that latter purpose, the LVMT has developed the LTK (**L**adta **T**ool**K**it), powerful computation implementation of LADTA main principles and associated solution methods (Aguiléra and Leurent, 2009).

LabDTA is implemented in TCL, a dynamic language that is commonly used for rapid prototyping and scripted applications. LabDTA is interfaced with the LTK. As the LTK is essentially a C API, this interface is rather useful to quickly set numerical experiments using the LTK.

LabDTA has notably been used to design the DUE algorithms proposed in this thesis. Several tests of the algorithm were conducted using LabDTA before actually implementing it in the LTK.

## 2 Overview of the toolbox

The toolbox is composed of 5 libraries of functions:

- `PWL_Function` is a set of procedures to deal piecewise linear (PWL) functions which are the basic variables of LabDTA,. Over 50 procedures are implemented, allowing basic arithemical operations on PWL functions (linear combination, multiplication, etc.), more complex algorithmic operations (mathematical programming and equation solving) and file input and output for different formats (including database files).

- `tt_models` provides functions to compute traffic propagation on the arc of a road network. Two models are proposed: the volume-delay model and the bottleneck model.

- `Network` implements a graph structure intended to represent dynamic transport networks. It is similar to the model presented in Chapter 1.

- `Network_Loading` implements the traffic loading algorithm presented in Appendix D.

- `DTC_Choice` implements the departure time choice algorithms that are presented in Chapter 8.

In addition to those libraries, two visualisation tools are provided with LabDTA:

- `SmallWin` allows to launch a small windows from a tcl script or shell and to display PWL functions dynamically. It is designed to tackle with an important number of graphs to display. Graphs can be exported to the eps format.

- `Netview` allows to display a network and to visualize functions associated to the nodes and the arcs.

**For more details about LabDTA, please contact the author.**

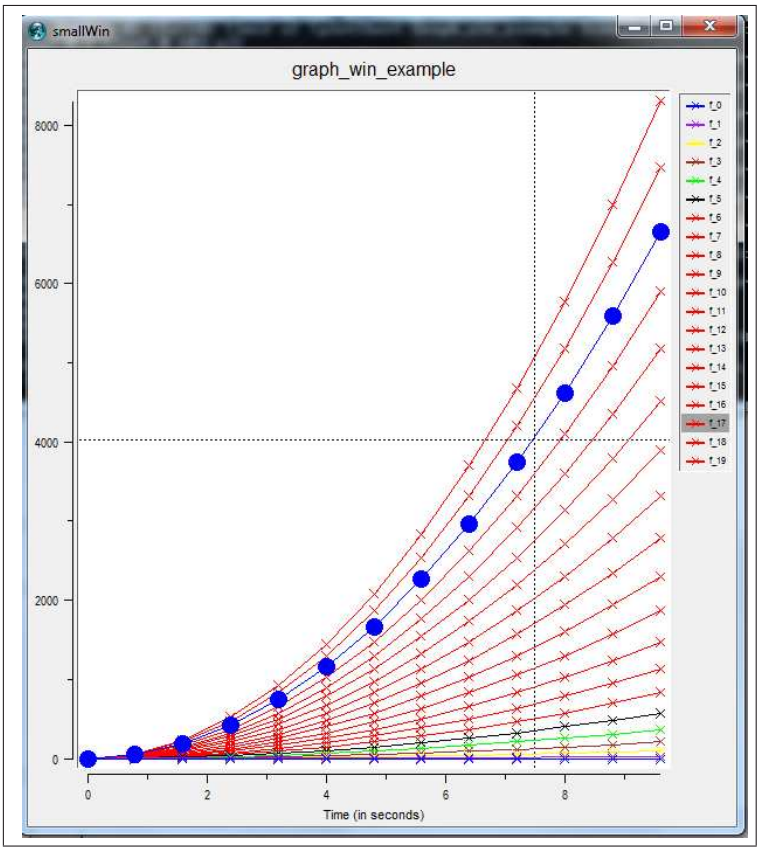



Figure E.1: SmallWin, a visualization tool for piecewise linear functions

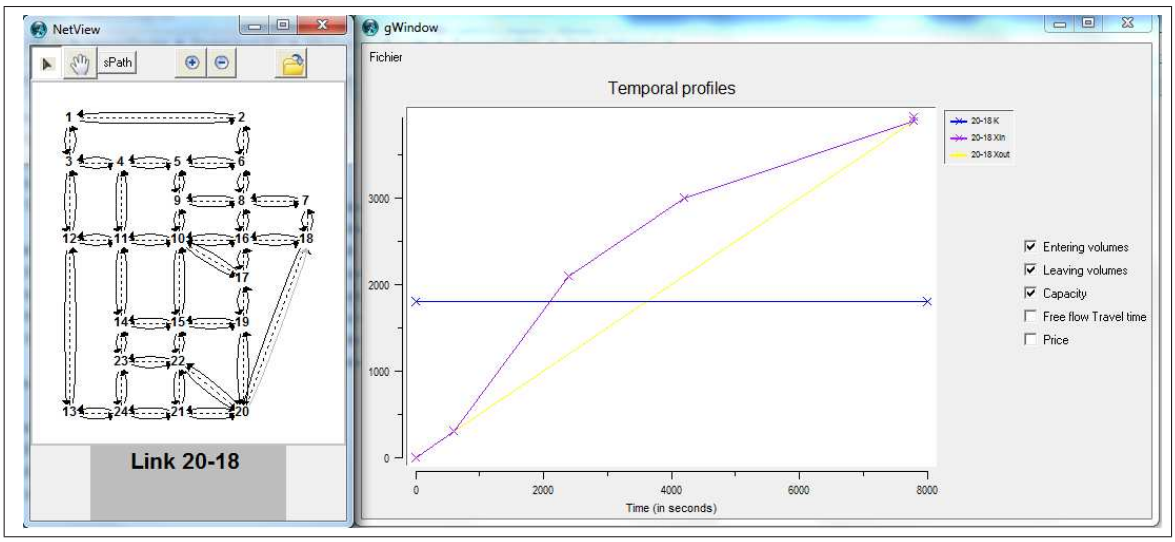



Figure E.2: Netview, a visualization tool for dynamic transport networks