

# Markovian traffic equilibrium\*

J-B. Baillon<sup>†</sup> and R. Cominetti<sup>‡</sup>

## Abstract

We analyze a stochastic dynamic programming model for traffic equilibrium on networks. In this model passengers move towards their destinations by a sequential process of arc selection based on a discrete choice model at every intermediate node in their trip. Route selection is therefore the outcome of a sequential process of arc choices while network flows correspond to the invariant measures of the corresponding Markov chains. The approach may handle different discrete choice models at every node, including the possibility of mixing deterministic and stochastic distribution rules. It can also be used over a multi-modal network in order to model the simultaneous selection of mode and route, as well as to treat the case of elastic demands. We establish the existence of a unique equilibrium, which is characterized as the solution of an unconstrained strictly convex minimization problem of low dimension. We report some numerical experiences comparing the performance of the method of successive averages and Newton's method on one small and one large network, and we prove the convergence of the former.

## 1 Introduction

Traffic assignment models describe the way in which the transportation demands flow through a given network. Users are supposed to behave rationally and to travel along the shortest available paths [17, Wardrop's principle]. However, because of congestion the path costs are themselves influenced by the flows and therefore the models give rise to equilibrium or fixed point problems.

Assignment models can be classified as *deterministic* or *stochastic* depending whether all users perceive the same costs or there is some variability among the population. In the first case the shortest paths connecting a given origin-destination pair (OD) are the same for all users and the flow distribution rule is deterministic, while in a stochastic setting the shortest paths depend on the particular user under consideration leading to a flow distribution principle based on random utility theory. Despite the similarity of both situations, the mathematical formulations found in the literature are rather different and both approaches remain somewhat disconnected. In this paper we describe a unifying framework for deterministic and stochastic traffic assignment, in the form of an unconstrained convex minimization problem that encompasses both types of assignment simultaneously. In this approach, the deterministic assignment becomes a particular instance of the stochastic assignment.

---

\*Partially supported by ECOS-Conicyt program under grant No. C00E05

<sup>†</sup>Université de Paris I.

<sup>‡</sup>Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático, Universidad de Chile. Partially supported by Fondap Matemáticas Aplicadas, CONICYT-Chile.

Stochastic route choice models can be traced back to Dial [6] where the demand on each OD pair is distributed among routes according to a logit discrete choice model. To reduce the computational burden of path enumeration, the choices are restricted to a smaller set of *efficient* routes, namely those for which each arc leads farther away from the origin and gets closer to the destination. This approach, further investigated in [16, Trahan] and [14, Sheffi and Powell], was conceived for the flow-independent case and assumes route costs to be independent Gumbel random variables. Since independence is an unlikely assumption when dealing with overlapping routes, Daganzo and Sheffi [5] proposed an alternative model based on a probit stochastic assignment which was solved by Montecarlo simulation. On a different direction, Dial's logit-based model was extended to the case of flow-dependent arc costs in [7, Fisk], where an equivalent optimization problem was found. The latter was formulated in the high dimensional space of route flows and the method of successive averages was proposed as a practical numerical scheme. Alternative methods based on a dual formulation and which do not require path enumeration were investigated in [11, Larrison, Liu and Patrickson]. Finally, we mention an attempt to extend the stochastic assignment to the case of transit networks in [12, Nguyen *et al.*]. For a more complete account on deterministic and stochastic traffic assignment we refer to the surveys of Florian and Hearn [8, 9].

A common feature of all the mentioned works is that each user on a given OD pair directly selects an optimal *route* by comparing it with all the available routes. From a modelling perspective this has several drawbacks. Firstly, partial information and limited discrimination capacity of the users suggest that the choices should be restricted to a few "*reasonable*" routes, but then defining a priori which routes should be considered is not obvious specially since travel times are flow-dependent. On the other hand, as already mentioned the logit-based models assume the independence of route costs which is unreasonable when dealing with overlapping routes, and also both logit and probit models distribute flows among all the available routes no matter how large their travel times are. Finally, models requiring path enumeration are computationally impractical for large networks.

In this paper we consider an alternative approach which is still based on discrete choice models but imbedded in a dynamic programming framework: route choice is no longer seen as an issue to be solved once and for all at the origin of each trip, but instead we consider it as the outcome of a sequential process of selection of arcs at every intermediate node. Passenger movements are then governed by an embedded Markov chain and therefore we call this type of assignment a *Markovian Traffic Equilibrium* or MTE. The idea is similar to the logit Markovian assignment of [2, Akamatsu] except that we consider general discrete choice models and at the same time we provide a compact dual characterization which seems more easily amenable to large scale computations. This approach does not require independence of the random route costs, and the number of alternative arcs to be considered at each node remains within the discrimination capabilities of the users with no need for arbitrarily reducing the set of possible routes. Moreover, by appropriately choosing the discrete choice models one can force the equilibrium assignments to use only nearly optimal routes and not every route as in the logit or probit route-based setting. Finally, the model avoids path enumeration and leads to computational procedures which are effective even for large networks.

An additional advantage of the MTE model is its flexibility to deal with different discrete choice models at every node, and even to mix deterministic and stochastic assignment rules within the same model. In particular, by considering multi-modal networks the approach applies directly to model

the simultaneous selection of mode and route. Also, the case of elastic demands may be treated as usual by adding *no-trip* arcs connecting directly each OD pair with travel cost equal to the inverse of the demand function.

The paper is structured as follows. Section §2 sets the notation and reviews the basics of deterministic traffic equilibrium, describing its equivalent formulation in terms of a primal and a dual optimization problems. Section §3 then introduces the MTE model and establishes an analog characterization which has exactly the same form as the deterministic dual characterization, providing a unified framework for both types of equilibrium. This dual program is unconstrained, strictly convex and coercive, implying the existence and uniqueness of an MTE. In the stochastic case the dual turns out to be smooth, and since it is also low dimensional (one variable per arc) it opens the way to solve large networks. Some extensions of the basic MTE model are briefly described in section §4, while section §5 reports some preliminary numerical results comparing the performance of the well known method of successive averages (which is interpreted as a variable metric gradient scheme) and Newton's method. These experiments show that the model is solvable even for large networks. We conclude the paper in §6 by giving sufficient conditions for the convergence of the method of successive averages. Appendix A at the end of the paper provides an analytical characterization of discrete choice models which plays an essential role in our analysis, while Appendix B computes the Hessians required to implement Newton's method.

## 2 Deterministic traffic equilibrium

Let  $G = (N, A)$  be a directed graph representing a traffic network, and  $D \subseteq N$  a set of destinations. Given the demands  $g_i^d \geq 0$  from each node  $i \neq d$  to each destination  $d \in D$ , a traffic assignment model seeks to predict how these demands flow throughout the network. More precisely, if we let  $R_i^d$  denote the set of simple paths connecting  $i$  to  $d$  (assumed nonempty) and writing  $R$  for the union of the  $R_i^d$ 's, the problem is to determine an "efficient" path-flow assignment  $h = (h_r)_{r \in R}$  with non-negative entries  $h_r \geq 0$  and satisfying the flow conservation constraints

$$g_i^d = \sum_{r \in R_i^d} h_r \quad \forall d \in D, i \neq d. \quad (1)$$

The efficiency of a path  $r \in R$  is measured by its generalized cost

$$c_r = \sum_{a \in r} t_a \quad (2)$$

where  $t_a = s_a(w_a)$  represents the *travel time* or *generalized cost* of arc  $a$  as a strictly increasing continuous function of the total flow on that arc, namely  $s_a : [0, \infty) \rightarrow [0, \infty)$  and

$$w_a = \sum_{r \ni a} h_r. \quad (3)$$

A deterministic model assumes that all users perceive the exact costs  $c_r$ , and then an efficient assignment is characterized by the fact that only shortest paths are used [17, Wardrop].

**Definition 1** Let  $H$  denote the set of all  $(w, h)$  satisfying (1) and (3) with  $h \geq 0$ . A pair  $(w, h) \in H$  is a deterministic traffic equilibrium iff for each destination  $d \in D$  and every  $i \neq d$  one has

$$(\forall r \in R_i^d) \quad h_r > 0 \Rightarrow c_r = \tau_i^d$$

where  $c_r$  is given by (2) and  $\tau_i^d = \min_{r \in R_i^d} c_r$ , with  $t_a = s_a(w_a)$ .

These equilibria are characterized [4, Beckman *et al.*] as the optimal solutions of the convex program

$$(P_H) \quad \min_{(w, h) \in H} \sum_{a \in A} \int_0^{w_a} s_a(z) dz.$$

Since  $H$  is compact there exist optimal solutions, while strict convexity implies that the optimal  $w^*$  is unique, although there may exist several path-flow assignments  $h$  which correspond to this  $w^*$ . We loosely refer to  $w^*$  as being “the” equilibrium flow.

## 2.1 Arc-flow formulation

As the number of paths is usually very large, the formulation  $(P_H)$  is computationally inefficient. An alternative formulation is obtained by considering variables  $v_a^d \geq 0$  representing the flow on arc  $a$  heading to destination  $d$ , and the set  $V$  of pairs  $(w, v) \geq 0$  satisfying the flow conservation constraints

$$g_i^d + \sum_{a \in A_i^-} v_a^d = \sum_{a \in A_i^+} v_a^d \quad \forall d \in D, i \neq d \quad (4)$$

as well as the total flow relations

$$w_a = \sum_{d \in D} v_a^d \quad \forall a \in A. \quad (5)$$

Clearly, to every feasible assignment  $(w, h) \in H$  it corresponds a unique pair  $(w, v) \in V$  by setting  $v_a^d = \sum \{h_r : a \in r, r \in R_i^d, i \neq d\}$ . However, not all elements in  $V$  are of this form since the latter allows flow along cycles which may be forbidden in  $H$ . In any case, the convex program

$$(P_V) \quad \min_{(w, v) \in V} \sum_{a \in A} \int_0^{w_a} s_a(z) dz$$

is a relaxation of  $(P_H)$  with the same optimal  $w^*$ . Moreover, any  $v$  with  $(w^*, v) \in V$  is optimal and the sets  $\{a \in A : v_a^d > 0\}$  are cycle-free so that  $v$  may be decomposed into path-flows  $h_r$  satisfying (1) and (3) (see *e.g.* [1, Theorem 2.1]). Any such decomposition gives an equilibrium and, conversely, every equilibrium is of this form.

## 2.2 Dual formulation

The main point when computing a traffic equilibrium is then to find the optimal arc-flow vector  $w^*$ . A dual problem which provides an alternative for this was described in [10, Fukushima]. We slightly re-formulate the latter in a form which is more convenient for comparison with the MTE model to be presented in the next section. For the sake of completeness we provide a short proof of the equivalence. The dual variables are the arc travel times  $t_a$  for  $a \in A$  and the time-to-destination variables  $\tau_i^d$  for  $d \in D$  and  $i \neq d$ . For notational convenience we set  $\tau_d^d = 0$  (considered as a constant rather than a constrained variable) and we let  $T$  denote the set of all vectors  $(t, \tau)$  satisfying

$$\tau_{i_a}^d \leq t_a + \tau_{j_a}^d \quad \forall a \in A, d \in D \quad (6)$$

where  $i_a$  and  $j_a$  are the initial and terminal nodes of arc  $a$ . We also extend the functions  $s_a(\cdot)$  to  $\mathbb{R}_-$  by setting  $s_a(w_a) = s_a(0) + w_a$  for  $w_a < 0$ .

**Theorem 1** *The convex program*

$$(D) \quad \min_{(t, \tau) \in T} \sum_{a \in A} \int_0^{t_a} s_a^{-1}(y) dy - \sum_{\substack{d \in D \\ i \neq d}} g_i^d \tau_i^d$$

has optimal solutions, and any such solution  $(t, \tau)$  satisfies  $w_a^* = s_a^{-1}(t_a)$ .

*Proof.* Let  $(t, \tau)$  be an optimal solution. Since the constraints of (D) are linear, there exist multipliers  $v_a^d \geq 0$  satisfying the Kuhn-Tucker conditions

$$\begin{aligned} (a) \quad & v_a^d [\tau_{i_a}^d - t_a - \tau_{j_a}^d] = 0 & \forall a \in A \\ (b) \quad & s_a^{-1}(t_a) = \sum_{d \in D} v_a^d & \forall a \in A \\ (c) \quad & g_i^d + \sum_{a \in A_i^-} v_a^d = \sum_{a \in A_i^+} v_a^d & \forall d \in D, i \neq d. \end{aligned}$$

Setting  $w_a = \sum_{d \in D} v_a^d$ , the pair  $(w, v)$  turns out to be feasible for  $(P_V)$ , while conditions (a), (b) and (6) imply that  $t_a$  and  $\tau_i^d$  are Lagrange multipliers for the constraints (5) and (4) respectively. Hence  $(w, v)$  is a stationary point for  $(P_V)$  and since the latter is a convex program it follows that  $(w, v)$  is optimal. Therefore  $w = w^*$  and the equality  $w_a^* = s_a^{-1}(t_a)$  results from (b).

The existence of optimal solutions for (D) follows by proceeding in the other direction: take  $(w^*, v)$  optimal for  $(P_V)$  and let  $t_a$  and  $\tau_i^d$  be corresponding multipliers. Then  $(t, \tau)$  is a stationary point for (D), hence an optimal solution.  $\blacksquare$

Notice that given any  $t$ , a corresponding optimal  $\tau = \bar{\tau}(t)$  may be found by solving shortest paths for each OD pair  $(i, d)$ , so that problem (D) is equivalent to the unconstrained, non-smooth, and strictly convex program

$$(\bar{D}) \quad \min_{t \in \mathbb{R}^{|A|}} \bar{\Phi}(t) \triangleq \sum_{a \in A} \int_0^{t_a} s_a^{-1}(y) dy - \sum_{\substack{d \in D \\ i \neq d}} g_i^d \bar{\tau}_i^d(t)$$

with unique optimal solution  $t_a^* = s_a(w_a^*)$ .

### 3 Markovian traffic equilibrium

A critical assumption in the deterministic model is the fact that all users perceive the same costs  $c_r$ . Now, the difficulty to discriminate paths with similar costs, as well as small variations in the perception of travel time among different users, suggest the necessity to relax this assumption. Stochastic models approach this question by assuming that users are randomly drawn from a large population having a variable perception of routes: the costs  $\tilde{c}_r$  become random variables and the demand  $g_i^d$  splits among the paths  $r \in R_i^d$  according to

$$h_r = g_i^d \mathbb{P}(\tilde{c}_r \leq \tilde{c}_p, \forall p \in R_i^d). \quad (7)$$

For instance, in the popular logit model the  $\tilde{c}_r$ 's are taken as independent Gumbel variables with expectation  $c_r$  and shape parameter  $\beta$ , leading to  $h_r = g_i^d e^{-\beta c_r} / \sum_{p \in R_i^d} e^{-\beta c_p}$ . Except for very small networks, this approach is impractical as it requires path enumeration. More importantly, the independence assumption is unrealistic when dealing with overlapping paths. For instance, in the small example below with 3 routes from  $i$  to  $d$ , all of them with the same expected cost  $c_r = 1$ , the logit rule assigns one third of the demand to each route. However, since both routes using the lower arc are almost identical the assignment  $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$  seems more natural. The latter is the solution obtained if we focus on *arc choices* rather than *path choices*: at node  $i$  there are only two arc options (upper and lower), both offering the same travel time so that one may reasonably expect that each one gets roughly one half of the demand. The half taking the lower arc faces a second choice at the intermediate node  $j$  where it splits again giving  $\frac{1}{4}$  of the demand on each of the two lower routes.

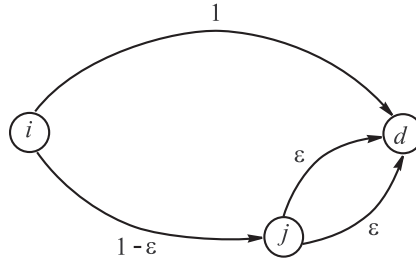


Figure 1: *path v/s arc choices on a small network*

In this paper we pursue the latter idea by looking at route selection as a recursive decision process based on arc choices. The variability within the population is introduced by modelling the cost of each arc as a continuous random variable  $\tilde{t}_a = t_a + \epsilon_a$  with  $\mathbb{E}(\epsilon_a) = 0$ . Consequently, the route costs  $\tilde{c}_r = \sum_{a \in r} \tilde{t}_a$ , and therefore the optimal costs  $\tilde{\tau}_i^d = \min_{r \in R_i^d} \tilde{c}_r$  as well as the cost  $\tilde{z}_a^d = \tilde{t}_a + \tilde{\tau}_{j_a}^d$  of an arc  $a$  in regard to destination  $d$ , become random variables which we write in the form  $\tilde{\tau}_i^d = \tau_i^d + \theta_i^d$  and  $\tilde{z}_a^d = z_a^d + \epsilon_a^d$  with  $\tau_i^d = \mathbb{E}(\tilde{\tau}_i^d)$  and  $z_a^d = \mathbb{E}(\tilde{z}_a^d)$ . Each user travelling towards  $d$  and reaching node  $i$  observes the variables  $\tilde{z}_a^d$  and then selects the arc  $a \in A_i^+$  with the lowest cost (the probability

of a tie is supposed to be zero). This process is repeated at each subsequent node giving rise to a recursive discrete choice model. Hence, for each destination  $d \in D$  we have an underlying finite-state Markov chain on the graph  $G$  with transition probabilities

$$P_{ij}^d = \begin{cases} \mathbb{P}(\tilde{z}_a^d \leq \tilde{z}_b^d, \forall b \in A_i^+) & \text{if } i = i_a, j = j_a \\ 0 & \text{otherwise.} \end{cases}$$

for  $i \neq d$ ; and  $P_{dd}^d = 1, P_{dj}^d = 0$  for  $j \neq d$  (*i.e.* destination  $d$  is an absorbing state of the chain). The distribution rule (7) is therefore replaced by a recursive scheme where the expected flow  $x_i^d$  entering a node  $i \neq d$  and directed towards  $d$  splits among the arcs  $a \in A_i^+$  according to

$$v_a^d = x_i^d \mathbb{P}(\tilde{z}_a^d \leq \tilde{z}_b^d, \forall b \in A_i^+). \quad (8)$$

Denoting  $\hat{P}^d = (P_{ij}^d)_{i,j \neq d}$  the restriction of the transition matrix to  $N \setminus \{d\}$  and letting  $g^d = (g_i^d)_{i \neq d}$ , the expected flow  $x^d = (x_i^d)_{i \neq d}$  for destination  $d$  is given by  $x^d = \sum_{k=0}^{\infty} [(\hat{P}^d)^k] g^d$ . This can also be expressed as  $x^d = g^d + (\hat{P}^d)' x^d$ , that is to say

$$x_i^d = g_i^d + \sum_{i \in A_i^-} v_a^d. \quad (9)$$

Before stating a formal definition, we make a short detour to show that all the relevant information for building a stochastic model is encapsulated in the functions

$$\varphi_i^d(z^d) \triangleq \mathbb{E}(\min_{a \in A_i^+} [z_a^d + \epsilon_a^d]),$$

which we take as the primary modelling objects. We denote by  $\mathcal{E}$  the class of functions that can be expressed in this form, and which are completely characterized in Appendix A at the end of the paper. For the moment it suffices to mention that these functions are componentwise non-decreasing, concave and smooth, and the transition probabilities may be expressed as

$$\mathbb{P}(\tilde{z}_a^d \leq \tilde{z}_b^d, \forall b \in A_i^+) = \frac{\partial \varphi_i^d}{\partial z_a^d}(z^d)$$

so that the flow distribution equations (8) and (9) can be restated as

$$\begin{cases} v_a^d = x_i^d \frac{\partial \varphi_i^d}{\partial z_a^d}(z^d) & \forall a \in A_i^+ \\ x_i^d = g_i^d + \sum_{a \in A_i^-} v_a^d. \end{cases} \quad (10)$$

On the other hand, since  $\tilde{z}_a^d = \tilde{t}_a + \tilde{\tau}_{j_a}^d$  we may write Bellman's equations of dynamic programming in the form  $\tilde{\tau}_i^d = \min_{a \in A_i^+} \tilde{z}_a^d$ , and by taking expectation we get

$$\begin{cases} z_a^d = t_a + \tau_{j_a}^d \\ \tau_i^d = \varphi_i^d(z^d) \end{cases}$$

which can be expressed solely in terms of the variables  $z_a^d$ 's as

$$z_a^d = t_a + \varphi_{j_a}^d(z^d) \quad (11)$$

or in terms of the variables  $\tau_i^d$  as

$$\tau_i^d = \varphi_i^d(t_a + \tau_{j_a}^d : a \in A). \quad (12)$$

We stress that (11) and (12) are totally equivalent: if  $z_a^d$  solves (11) then  $\tau_i^d \triangleq \varphi_i^d(z^d)$  is a solution for (12), and conversely if  $\tau_i^d$  solves (12) then  $z_a^d \triangleq t_a + \tau_{j_a}^d$  solves (11). With these preliminary remarks we proceed to formalize the notion of stochastic equilibrium. Throughout we assume that

$$(H_0) \left\{ \begin{array}{l} \text{we are given a family } \varphi_i^d \in \mathcal{E} \text{ with } \varphi_d^d \equiv 0 \text{ and strictly increasing continuous travel time} \\ \text{functions } s_a : \mathbb{R}_+ \rightarrow \mathbb{R} \text{ with } \lim_{w_a \uparrow \infty} s_a(w_a) = \infty, t_a^0 = s_a(0) \geq 0 \text{ and } \varphi_i^d(t^0) > 0 \text{ for all } i \neq d \end{array} \right.$$

and as before we extend  $s_a(\cdot)$  to  $\mathbb{R}_-$  by setting  $s_a(w_a) = s_a(0) + w_a$  for  $w_a < 0$ .

**Definition 2** A pair  $(w, v) \geq 0$  is a Markovian Traffic Equilibrium (MTE) iff  $w_a = \sum_{d \in D} v_a^d$  where the  $v_a^d$ 's satisfy the flow distribution equations (10) with  $z^d$  solving (11) for  $t_a = s_a(w_a)$ .

Although this notion looks rather different from the deterministic one in Definition 1, we will show that they are in fact closely related. Before we proceed, let us check that the equations (10)-(12) uniquely define  $v, x, z$  and  $\tau$  as implicit functions of  $t$ . In order to fix the notations used in the sequel we set  $\tau^d = (\tau_i^d)_{i \neq d}$ ,  $x^d = (x_i^d)_{i \neq d}$ ,  $v^d = (v_a^d)_{a \in A}$ ,  $g^d = (g_i^d)_{i \neq d}$ , and we consider the matrices  $\hat{P}^d(z^d) = (P_{ij}^d)_{i,j \neq d}$  and  $\hat{Q}^d(z^d) = (Q_{ia}^d)_{i \neq d, a \in A}$  with entries

$$P_{ij}^d = \begin{cases} \frac{\partial \varphi_i^d}{\partial z_a^d}(z^d) & \text{if } i = i_a \text{ and } j = j_a \\ 0 & \text{otherwise} \end{cases}$$

$$Q_{ia}^d = \begin{cases} \frac{\partial \varphi_i^d}{\partial z_a^d}(z^d) & \text{if } i = i_a \\ 0 & \text{otherwise.} \end{cases}$$

We begin by stating a technical Lemma.

**Lemma 1** Assume  $(H_0)$  and let  $t = (t_a)_{a \in A}$  with  $t_a \geq t_a^0$ . If  $(z^d, \tau^d)$  solves (11)-(12) then

(a) for each  $i \neq d$  there is  $j \in N$  with  $P_{ij}^d > 0$  and  $\tau_j^d < \tau_i^d$ ,

(b) the matrix  $[I - \hat{P}^d(z^d)]$  is invertible,

(c) equation (10) has a unique solution  $v^d = \hat{Q}^d(z^d)' x^d$  with  $x^d = [I - \hat{P}^d(z^d)]^{-1} g^d$ .

*Proof.* (a) Let us set  $\hat{z}_a^d = t_a + \tau_i^d$  for all  $a \in A$ . Since  $\varphi_i^d$  is concave and smooth we have

$$\begin{aligned} \varphi_i^d(\hat{z}^d) &\leq \varphi_i^d(z^d) + \sum_{a \in A} \frac{\partial \varphi_i^d}{\partial z_a^d}(z^d) (\hat{z}_a^d - z_a^d) \\ &= \tau_i^d + \sum_{a \in A_i^+} P_{ij_a}^d (\tau_i^d - \tau_{j_a}^d). \end{aligned}$$

Now, Proposition A.2(b) in the appendix gives  $\varphi_i^d(\hat{z}^d) = \tau_i^d + \varphi_i^d(t)$ , while assumption  $(H_0)$  implies  $\varphi_i^d(t) \geq \varphi_i^d(t^0) > 0$ , so that

$$0 < \sum_{a \in A_i^+} P_{j_a}^d (\tau_i^d - \tau_{j_a}^d)$$

from which (a) follows at once.

(b) Using (a) inductively, for all  $i \neq d$  we may find a sequence of nodes  $i_0, i_1, \dots, i_m$  with  $i_0 = i, i_m = d$  and  $P_{i_k i_{k+1}}^d > 0$ . Hence the Markov chain started from  $i$  has a positive probability of reaching the absorbing state  $d$ , so that  $\hat{P}^d(z^d)^k$  is strictly sub-markovian for  $k$  large enough and then  $[I - \hat{P}^d(z^d)]$  is invertible.

(c) The first equation of (10) gives  $v^d = \hat{Q}^d(z^d)'x^d$ , which substituted into the second equation yields the linear system  $x^d = g^d + \hat{P}^d(z^d)'x^d$  so that  $x^d = [I - \hat{P}^d(z^d)']^{-1}g^d$ . ■

**Proposition 1** *Assume  $(H_0)$  and let  $t = (t_a)_{a \in A}$  with  $t_a \geq t_a^0$ . Then, for each destination  $d \in D$ , the equations (10)-(12) have unique solutions  $v^d = v^d(t)$ ,  $x^d = x^d(t)$ ,  $z^d = z^d(t)$  and  $\tau^d = \tau^d(t)$ . Moreover, the functions  $t \mapsto \tau_i^d(t)$  and  $t \mapsto z_a^d(t)$  are concave and smooth.*

*Proof.* According to Lemma 1(c) and since  $z_a^d = t_a + \tau_{j_a}^d$ , it suffices to analyze  $\tau^d(t)$ . Let us then examine (12) for a fixed  $d$ .

EXISTENCE. Consider the sequence generated inductively by  $\tau_i^{k+1} = \varphi_i^d(t_a + \tau_{j_a}^k)$  started from  $\tau_i^0 = 0$ . Since the  $\varphi_i^d$ 's are componentwise non-decreasing, using  $(H_0)$  we get  $\tau^1 \geq \tau^0$  and then it follows inductively that  $\tau^{k+1} \geq \tau^k$ . Denoting  $\bar{\tau}_i$  the shortest distance from  $i$  to  $d$ , Proposition A.1 in the appendix yields

$$\varphi_i^d(t_a + \bar{\tau}_{j_a}) \leq \min_{a \in A_i^+} [t_a + \bar{\tau}_{j_a}] = \bar{\tau}_i$$

and since we obviously have  $\tau^0 \leq \bar{\tau}$ , it follows inductively that  $\tau^k \leq \bar{\tau}$  for all  $k$ . Therefore the non-decreasing sequences  $\tau_i^k$  have a limit  $\tau_i^d$  satisfying  $\tau_i^d = \varphi_i^d(t_a + \tau_{j_a}^d)$ , *i.e.* a solution of (12).

UNIQUENESS. Let  $\tau^1, \tau^2$  be 2 solutions, and denote  $\alpha = \max_{i \in N} [\tau_i^1 - \tau_i^2]$  and  $N^*$  the set of nodes  $i \in N$  where this maximum is attained. Since  $\tau_{j_a}^1 \leq \tau_{j_a}^2 + \alpha$ , for all  $i \in N^*$  we have

$$\begin{aligned} \tau_i^1 = \varphi_i^d(t_a + \tau_{j_a}^1) &\leq \varphi_i^d(t_a + \tau_{j_a}^2 + \alpha) \\ &= \varphi_i^d(t_a + \tau_{j_a}^2) + \alpha \\ &= \tau_i^2 + \alpha \\ &= \tau_i^1. \end{aligned}$$

A contradiction argument based on strict monotonicity then shows that whenever  $\frac{\partial \varphi_i^d}{\partial z_a^d}(t_a + \tau_{j_a}^1) > 0$  one must also have  $\tau_{j_a}^1 = \tau_{j_a}^2 + \alpha$ , that is  $j_a \in N^*$ . Hence, using Lemma 1(a) we deduce that for all  $i \in N^* \setminus \{d\}$  we may find  $j \in N^*$  with  $\tau_j^1 < \tau_j^1$ . Proceeding inductively we eventually reach  $j = d$  proving that  $d \in N^*$  and therefore  $\alpha = \tau_d^1 - \tau_d^2 = 0$ . It follows that  $\tau^1 \leq \tau^2$  and exchanging the roles we obtain  $\tau^1 = \tau^2$  proving uniqueness.

CONCAVITY. Let  $\tau^1 = \tau(t^1)$  and  $\tau^2 = \tau(t^2)$ , and set  $t^\alpha = \alpha t^1 + (1-\alpha)t^2$  and  $\tau^\alpha = \alpha\tau^1 + (1-\alpha)\tau^2$  with  $\alpha \in (0, 1)$ . Since the functions  $\varphi_i^d$  are concave, we obtain

$$\begin{aligned}\varphi_i^d(t_a^\alpha + \tau_{j_a}^\alpha) &\geq \alpha\varphi_i^d(t_a^1 + \tau_{j_a}^1) + (1-\alpha)\varphi_i^d(t_a^2 + \tau_{j_a}^2) \\ &= \alpha\tau_i^1 + (1-\alpha)\tau_i^2 \\ &= \tau_i^\alpha.\end{aligned}$$

Hence, the recursion  $\tau_i^{k+1} = \varphi_i^d(t_a^\alpha + \tau_{j_a}^k)$  started from  $\tau^0 = \tau^\alpha$  generates a non-decreasing sequence. Moreover, proceeding as in the existence part of the proof we may show inductively that  $\tau_i^k \leq \bar{\tau}_i^\alpha$  with  $\bar{\tau}_i^\alpha$  the shortest distance from  $i$  to  $d$  with arc travel times  $t_a^\alpha$ . Hence  $\tau^k$  converges and by continuity the limit is a solution of (12) so that  $\tau^k \rightarrow \tau(t^\alpha)$ . Moreover, since  $\tau^k$  is non-decreasing this limit is greater than  $\tau^0$ , that is to say

$$\tau_i^d(t^\alpha) \geq \tau_i^0 = \alpha\tau_i^d(t^1) + (1-\alpha)\tau_i^d(t^2)$$

which proves that  $\tau_i^d(\cdot)$  is concave.

SMOOTHNESS. This is a direct consequence of the Implicit Function Theorem. Indeed, noting that  $\tau_d^d = 0$  we may reduce (12) to a system in the variables  $(\tau_i^d)_{i \neq d}$ . The Jacobian of this reduced system is  $[I - \hat{P}^d]$  which is invertible according to Lemma 1(b), so the conclusion follows. ■

To simplify the notation, in the sequel we denote  $\hat{P}^d(t) = \hat{P}^d(z^d(t))$  and  $\hat{Q}^d(t) = \hat{Q}^d(z^d(t))$ , so that  $x^d(t) = [I - \hat{P}^d(t)]^{-1}g^d$  and  $v^d(t) = \hat{Q}^d(t)'x^d(t)$ .

**Corollary 1** *We have  $\frac{\partial \tau^d}{\partial t} = [I - \hat{P}^d(t)]^{-1}\hat{Q}^d(t)$ . Moreover, the function  $\psi^d(t) = \sum_{i \neq d} g_i^d \tau_i^d(t)$  is concave and smooth with  $\nabla \psi^d(t) = v^d(t)$ .*

*Proof.* An implicit differentiation of (12) yields  $\frac{\partial \tau^d}{\partial t} = \hat{Q}^d(t) + \hat{P}^d(t) \frac{\partial \tau^d}{\partial t}$ , from which we easily get the expression for  $\frac{\partial \tau^d}{\partial t}$ . Then, using the formula of  $\psi^d(t)$  we deduce

$$\nabla \psi^d(t) = \left( \frac{\partial \tau^d}{\partial t} \right)' g^d = \hat{Q}^d(t)' [I - \hat{P}^d(t)]^{-1} g^d = \hat{Q}^d(t)' x^d(t) = v^d(t). \quad \blacksquare$$

We are ready to establish our main result for the stochastic traffic equilibrium. We will prove that, with the new meaning of the functions  $\tau_i^d(\cdot)$ , the characterization  $(\bar{D})$  of a deterministic assignment remains valid for the stochastic assignment. This provides a unified framework for both approaches.

**Theorem 2** *Under  $(H_0)$  there exists a unique MTE which is given by*

$$(C) \quad \begin{cases} w_a = s_a^{-1}(t_a^*) \\ v_a^d = v_a^d(t_a^*) \end{cases}$$

where  $t^*$  denotes the unique optimal solution of the smooth strictly convex program

$$(S) \quad \min_{t \in \mathbb{R}^{|A|}} \Phi(t) \triangleq \sum_{a \in A} \int_0^{t_a} s_a^{-1}(z) dz - \sum_{\substack{d \in D \\ i \neq d}} g_i^d \tau_i^d(t).$$

*Proof.* From Definition 2 and Proposition 1 it follows that a pair  $(w, v) \geq 0$  is a stochastic equilibrium if and only if  $w_a = \sum_{d \in D} v_a^d$  with  $v^d = v^d(t)$  and  $t_a = s_a(w_a)$ . This is equivalent to say that  $(w, v)$  may be written as  $w_a = s_a^{-1}(t_a)$  and  $v^d = v^d(t)$  with  $t$  satisfying  $s_a^{-1}(t_a) = \sum_{d \in D} v_a^d(t)$ . Now, using Corollary 1 we observe that the latter corresponds to  $\nabla \Phi(t) = 0$ , and therefore  $(w, v)$  is a stochastic equilibrium iff it is of the form (C) with  $t^*$  an optimal solution of (S). To conclude we observe that since the  $\tau_i^d(\cdot)$ 's are concave and the  $s_a^{-1}(\cdot)$ 's are strictly increasing with  $s_a^{-1}(t_a) \rightarrow \pm\infty$  when  $t_a \rightarrow \pm\infty$ , then  $\Phi(\cdot)$  is strictly convex and coercive so that (S) has a unique optimal solution  $t^*$ . ■

We stress the fact that (S) is a smooth and strictly convex unconstrained problem, and the dimension of the space  $\mathbb{R}^{|A|}$  is very low compared to the models based on route flows. These nice features must be weighted against the difficulty in computing the implicit functions  $\tau_i^d(\cdot)$ . Notice however that the existence proof in Proposition 1 provides a constructive algorithm for this, which will be exploited in section §5 to obtain numerical schemes for computing the MTE. Alternatively, introducing variables  $\tau_i^d$  and  $z_a^d$  one may rewrite (S) as a constrained program

$$(\tilde{S}) \quad \begin{cases} \min_{(t, \tau, z)} \sum_{a \in A} \int_0^{t_a} s_a^{-1}(z) dz - \sum_{\substack{d \in D \\ i \neq d}} g_i^d \tau_i^d \\ \text{s.t.} \\ \tau_i^d = \varphi_i^d(z^d) \\ z_a^d = t_a + \tau_{j_a}^d \end{cases}$$

The latter may be dualized to obtain a *primal* characterization of MTE which is the analog of the primal characterization  $(P_V)$  for the deterministic equilibrium, namely

**Theorem 3** *The MTE is the unique optimal solution of the problem*

$$(P) \quad \min_{(w, v) \in V} \sum_{a \in A} \int_0^{w_a} s_a(z) dz + \sum_{d \in D} \chi^d(v^d)$$

where  $\chi^d(v^d) = \sup_{z^d} \sum_{a \in A} (\varphi_{i_a}^d(z^d) - z_a^d) v_a^d$  is positively homogeneous and convex.

*Proof.* Let us check that the stationarity conditions for (P) correspond to the conditions for MTE. We begin by noting that the constraints  $(w, v) \geq 0$  are implicit in the flow conservation constraints and the fact that  $\chi^d(v^d) = \infty$  whenever  $v_a^d < 0$  for some  $a \in A$ . Hence, we may consider (4) and (5) as the only constraints in (P) so that taking multipliers  $\tau_i^d$  and  $t_a$  and forming the lagrangian

$$\mathcal{L} = \sum_{a \in A} \int_0^{w_a} s_a(z) dz + \sum_{d \in D} \chi^d(v^d) + \sum_{\substack{d \in D \\ i \neq d}} \tau_i^d \left[ \sum_{A_i^+} v_a^d - g_i^d - \sum_{A_i^-} v_a^d \right] + \sum_{a \in A} t_a \left[ \sum_{d \in D} v_a^d - w_a \right]$$

stationarity amounts to the conditions  $s_a(w_a) = t_a$  and  $v^d \in \partial \chi^d(v^d)$  where  $v_a^d = \tau_{i_a}^d - t_a - \tau_{j_a}^d$  (with  $\tau_d^d = 0$ ). By subdifferential calculus the latter means that there exists  $z^d$  an optimal solution of  $\chi^d(v^d)$  such that  $\varphi_{i_a}^d(z^d) - z_a^d = v_a^d$ . From Proposition A.2(b) we observe that  $(z_a^d : a \in A_i^+)$  may be modified by a constant without affecting its optimality, and therefore we may assume with no loss of generality that  $\varphi_i^d(z^d) = \tau_i^d$ . The equality  $\varphi_{i_a}^d(z^d) - z_a^d = \tau_{i_a}^d - t_a - \tau_{j_a}^d$  then becomes  $z_a^d = t_a + \tau_{j_a}^d$  so

that the subdifferential condition is equivalent to (11)-(12). Finally, we observe that the optimality of  $z^d$  is equivalent to

$$v_a^d = \frac{\partial \varphi_{ia}^d}{\partial z_a^d} \sum_{a \in A_i^+} v_a^d$$

which corresponds to (10). Therefore the stationarity conditions for (P) correspond to (10) and (11) with  $t_a = s_a(w_a)$ , which are precisely the conditions that define the MTE. ■

We observe that  $\chi^d(v^d) = \sum_{i \neq d} \chi_i^d(v^d)$  with

$$\chi_i^d(v^d) = \sup_{z^d} \sum_{a \in A_i^+} (\varphi_i^d(z^d) - z_a^d) v_a^d.$$

In particular, in the deterministic case where we have  $\varphi_i^d(z^d) = \min_{a \in A_i^+} z_a^d$  it follows that  $\chi_i^d(v^d) \equiv 0$  and therefore (P) reduces exactly to  $(P_V)$ . Also, in the case of a logit Markov model where

$$\varphi_i^d(z^d) = -\frac{1}{\beta_i^d} \ln(\sum_{a \in A_i^+} e^{-\beta_i^d z_a^d})$$

a straightforward computation yields

$$\chi_i^d(v^d) = \frac{1}{\beta_i^d} \left[ \sum_{a \in A_i^+} v_a^d \ln(v_a^d) - (\sum_{a \in A_i^+} v_a^d) \ln(\sum_{a \in A_i^+} v_a^d) \right]$$

so that (P) coincides with Akamatsu's logit Markovian model [2] where the case  $\beta_i^d \equiv \theta$  is considered.

## 4 Extensions

### 4.1 Arc capacities and saturation

So far, the volume-delay functions  $s_a(\cdot)$  were assumed to be defined over  $[0, \infty)$ . Since very often the arcs have a maximal saturation capacity, it is useful to extend the model by considering strictly increasing continuous functions  $s_a : [0, \bar{w}_a) \rightarrow [0, \infty)$ . The saturation capacity  $\bar{w}_a$  is supposed strictly positive and may be infinite for those arcs which are not subject to saturation. To ensure feasibility we must assume that these capacities are large enough for the given demands, namely

**Theorem 4** *Assume  $(H_0)$  with  $s_a : [0, \bar{w}_a) \rightarrow [0, \infty)$  such that  $\lim_{w_a \rightarrow \bar{w}_a} s_a(w_a) = \infty$ , and suppose that there is a vector  $(\hat{w}, \hat{v})$  satisfying (4) and (5) with  $\hat{w}_a < \bar{w}_a$  for all  $a \in A$ . Then the conclusions of Theorem 2 remain valid.*

*Proof.* It suffices to establish the coercivity of the objective function  $\Phi(t)$ , for which we prove that the recession function  $\Phi^\infty(t) \triangleq \lim_{\lambda \rightarrow \infty} \frac{\Phi(\lambda t)}{\lambda}$  satisfies  $\Phi^\infty(t) > 0$  for all  $t \neq 0$ . The recession function of  $\sum_{a \in A} \int_0^{t_a} s_a^{-1}(z) dz$  is equal to  $\sum_{a \in A} \bar{w}_a t_a$  if  $t \geq 0$  and  $+\infty$  otherwise. In order to compute the recession functions of the  $\tau_i^d(\cdot)$ 's we exploit equation (12) to obtain

$$(\tau_i^d)^\infty(t) = (\varphi_i^d)^\infty(t_a + (\tau_{j_a}^d)^\infty(t)). \quad (13)$$

Now, the definition of the class  $\mathcal{E}$  and Lebesgue's theorem imply that  $(\varphi_i^d)^\infty(z^d) = \min\{z_a^d : a \in A_i^+\}$ , and therefore (13) shows that  $(\tau_i^d)^\infty(t)$  is the shortest distance from  $i$  to  $d$ , namely  $(\tau_i^d)^\infty(t) = \bar{\tau}_i^d(t)$ . Combining these facts we deduce that

$$\Phi^\infty(t) = \begin{cases} \sum_{a \in A} \bar{w}_a t_a - \sum_{\substack{d \in D \\ i \neq d}} g_i^d \bar{\tau}_i^d(t) & \text{if } t \geq 0 \\ +\infty & \text{otherwise.} \end{cases}$$

Multiplying the inequalities  $\bar{\tau}_{i_a}^d \leq t_a + \bar{\tau}_{j_a}^d$  by  $\hat{v}_a^d$  and adding over all the arcs  $a \notin A_d^+$  we get  $\sum_{i \neq d} g_i^d \bar{\tau}_i^d \leq \sum_{a \in A} t_a \hat{v}_a^d$ . Then, summing over all  $d \in D$  and assuming that  $t \geq 0, t \neq 0$  we obtain

$$\sum_{\substack{d \in D \\ i \neq d}} g_i^d \bar{\tau}_i^d(t) \leq \sum_{a \in A} t_a \hat{w}_a < \sum_{a \in A} t_a \bar{w}_a$$

which implies  $\Phi^\infty(t) > 0$  as was to be proved. ■

## 4.2 Mixed deterministic/stochastic assignment

The functions  $\varphi_i^d \in \mathcal{E}$  in the stochastic model are used to describe the flow distribution rule (10), and may differ from node to node: at some nodes one could consider logit distribution, while other nodes may be governed by probit or other discrete choice models. On the other hand, the deterministic model assumes that the flow entering each node is distributed among optimal arcs, which may also be written in the form (10) by taking  $\varphi_i^d(z^d) = \max\{z_a^d : a \in A_i^+\}$  and replacing the gradient by the subdifferential. This further explains the analogy between the characterizations  $(\bar{D})$  and  $(S)$ , and leads naturally to consider the possibility of a hybrid model where some nodes have a stochastic distribution rule while other nodes are deterministic. The analysis carries over to this more general setting with  $(S)$  characterizing the traffic equilibrium, though  $\Phi(\cdot)$  will no longer be smooth.

## 4.3 Simultaneous mode/route selection and elastic demands.

Noting that the graph  $G = (N, A)$  need not be limited to a single mode, the previous framework turns out to be flexible enough to handle more complex decision processes such as the simultaneous choice of *mode* and *route*. To this end it suffices to apply the model over a multi-modal graph built by connecting every origin and destination to one or several corresponding nodes on the subgraphs representing the basic transportation modes (car, bus, metro, walk, etc.). Combined modes such as car-metro-walk may be easily included by connecting the nodes on the corresponding subgraphs through additional *transfer arcs*. At every origin node one may adopt a particular distribution rule based on a logit or probit model, while at other nodes (*e.g.* the metro sub-network) one may use a deterministic rule. A further extension concerns the modelling of elastic demands. The option of not making a trip for a given OD pair may be simulated as usual by adding a no-trip arc which connects directly the origin to the destination, with cost equal to the inverse of the demand function (see [8]).

## 5 Numerical experiments

In this short section we describe some numerical tests for solving the minimization problem (S). According to Corollary 1 the derivatives of the objective function  $\Phi(t)$  in problem (S) are given by

$$\frac{\partial \Phi}{\partial t_a} = s_a^{-1}(t_a) - \tilde{w}_a(t)$$

with  $\tilde{w}_a(t) = \sum_{d \in D} v_a^d(t)$ . The computation of these derivatives requires solving first the system (12) in order to find the functions  $\tau_i^d(t)$  and then solving (10) to find  $v_a^d(t)$ . As seen from the existence proof of Proposition 1 this may be done for each destination  $d \in D$  by a fixed point iteration

$$FP(t) \quad \begin{cases} - \text{Iterate } \tau_i^{d,n+1} = \varphi_i^d(t_a + \tau_{j_a}^{d,n}) \text{ to find an estimate } \tau_i^d \sim \tau_i^d(t) \\ - \text{Compute } \frac{\partial \varphi_i^d}{\partial z_a^d}(z^d) \text{ with } z_a^d = t_a + \tau_{j_a}^d \text{ and build the matrices } \hat{P}^d \text{ and } \hat{Q}^d \\ - \text{Compute } x^d = [I - (\hat{P}^d)']^{-1} g^d \text{ and } v^d = (\hat{Q}^d)' x^d \\ - \text{Aggregate } \tilde{w} = \sum_{d \in D} v^d \end{cases}$$

A gradient method would use the above to estimate  $\tilde{w}^k \sim \tilde{w}(t^k)$  and then use the update

$$(G_t) \quad t_a^{k+1} = t_a^k - \alpha_k [s_a^{-1}(t_a^k) - \tilde{w}_a^k]$$

with  $\alpha_k$  a suitably chosen stepsize. Since performing a linesearch is too expensive, a normalized gradient stepsize was implemented by taking  $\alpha_k = \lambda_k / \|h^k\|$  with  $h_a^k = [s_a^{-1}(t_a^k) - \tilde{w}_a^k]$  and  $\lambda_k > 0$  such that  $\lambda_k \rightarrow 0$  and  $\sum_k \lambda_k = \infty$ . Although [15, Theorem 2.3] guarantees that  $t^k \rightarrow t^*$ , in our numerical tests the convergence was extremely slow (as usual for a gradient method).

An alternative method is obtained by using the change of variables  $t_a = s_a(w_a)$  to transform (S) into a strictly convex program in the total flow variables  $w_a$ , that is to say, minimizing the function  $\Psi(w) = \Phi(s(w))$ . The derivatives of the latter are given by  $\frac{\partial \Psi}{\partial w_a} = [w_a - \tilde{w}(s(w))] s_a'(w_a)$  so that the gradient iteration now becomes

$$(G_w) \quad w_a^{k+1} = w_a^k - \alpha_k [w_a^k - \tilde{w}_a^k] s_a'(w_a^k).$$

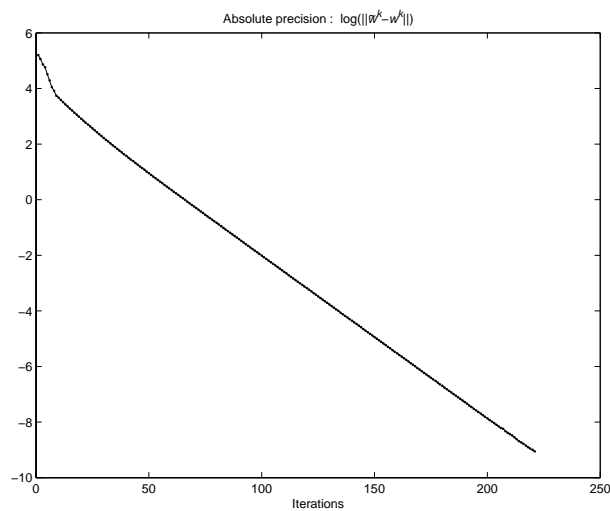
This was also implemented with a normalized stepsize but the convergence was again too slow. A more efficient variant is the method of successive averages which slightly modifies the latter as

$$(MSA) \quad w_a^{k+1} = w_a^k - \alpha_k [w_a^k - \tilde{w}_a^k]$$

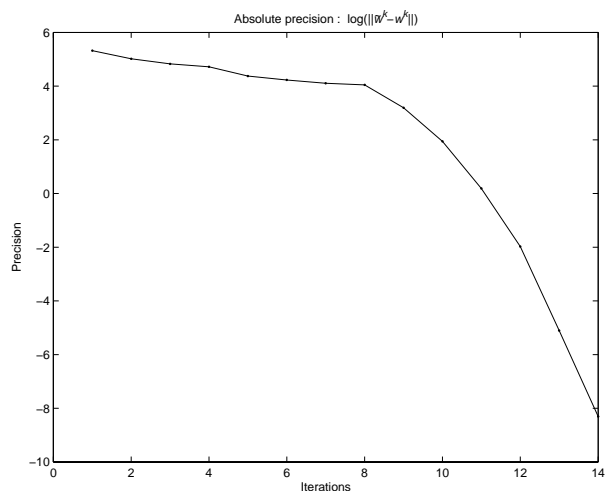
and which may be interpreted as a variable metric gradient method.

A Matlab implementation of MSA with constant stepsize  $\alpha_k \equiv \alpha$  was tested on the traffic network of Siouxfalls, a small network of 24 nodes and 76 arcs with travel times  $s_a(w_a) = t_a^0 [1 + b_a (\frac{w_a}{c_a})^{p_a}]$ . A recursive logit discrete choice model was considered by taking  $\varphi_i^d(z^d) = -\ln[\sum_{a \in A_i^+} \exp(-\beta_i^d z_a^d)] / \beta_i^d$ . Figure 2 plots the precision  $\log(\|\tilde{w}^k - w^k\|)$  along the iterations showing a linear rate of convergence. The method attains high absolute accuracy<sup>1</sup> but the number of iterations is large. However, the cost per iteration is low and the overall CPU time on a 1.6Mhz processor was 3.7 [s].

<sup>1</sup>The relative accuracy  $\|\tilde{w}^k - w^k\| / \|w^k\|$  is even higher reaching  $10^{-14}$ .

Figure 2: *MSA iterations — Siouxfalls*

In order to speed up convergence, the MSA iteration was combined with a Newton method which was activated after MSA reached a 10% relative precision, *i.e.*  $\|\tilde{w}^k - w^k\| \leq 0.1\|w^k\|$ . Figure 3 illustrates the performance of this variant. The faster convergence compensates the additional work involved in computing Newton's direction, with a significant reduction in CPU time to 0.7 [s].

Figure 3: *MSA-Newton iterations — Siouxfalls*

Additional tests were carried out on the larger network of Chicago which comprises 2950 arcs and 933 nodes (387 of which are destinations). The performance of both methods was very similar to the case of Siouxfalls reaching a precision  $\|\tilde{w}^k - w^k\| \leq 10^{-9}$  after 234 iterations and 29 [min] of CPU for the case of MSA, and 14 iterations and 11 [min] of CPU for Newton. In the latter case the Newton direction was computed by a conjugate gradient method that does not require explicitly

the Hessian (whose computation becomes expensive for large networks, see Appendix B) but only to evaluate products of this Hessian by vectors which can be done much more efficiently. The advantage of Newton's method over MSA is less clear for medium precision computation. For instance, MSA takes 6.3 [min] to reach a 0.1% relative accuracy, while Newton requires 4.3 [min] for this precision.

## 6 Convergence of MSA

We conclude the paper by establishing a sufficient condition for the convergence of MSA. This result can be seen as a discrete analog of the convergence analysis for continuous Riemannian gradient flows presented in [3, Alvarez, Bolte and Brahic].

**Theorem 5** *Assume  $(H_0)$  with  $s_a : [0, \infty) \rightarrow [0, \infty)$  of class  $C^2$ . If  $\lambda_k > 0$  is such that  $\sum_k \lambda_k = \infty$  and  $\sum_k \lambda_k^2 < \infty$ , then the sequence  $(w_k)_{k \in \mathbb{N}}$  generated by the MSA iteration  $w^{k+1} = (1 - \lambda_k)w^k + \lambda_k \tilde{w}^k$  converges to the MTE  $w^*$ .*

*Proof.* We begin by noting that  $\tilde{w}^k$  satisfies flow conservation so that the sequences  $\tilde{w}^k$  and  $w^k$  remain bounded. Proceeding as in §5 we rewrite MSA as the variable metric subgradient scheme

$$\frac{w^{k+1} - w^k}{\lambda_k} \in -D(w^k)^{-1} \partial \Psi(w^k)$$

where  $D(w) = \text{diag}[s'_a(w_a) : a \in A]$ . The convexity inequality then gives

$$\Psi(w^k) + \langle D(w^k) \frac{w^{k+1} - w^k}{\lambda_k}, w^k - w^* \rangle \leq \Psi(w^*)$$

which we may rewrite as

$$[\Psi(w^k) - \Psi(w^*)] \lambda_k + \langle \nabla h(w^k), w^{k+1} - w^k \rangle \leq 0 \quad (14)$$

with  $h(w) = \sum_{a \in A} \int_{w_a^*}^{w_a} s'_a(z) (z - w_a^*) dz$ . We notice that  $h(w^*) = 0$  and  $h(w) > 0$  for  $w \neq w^*$ . Now, by the mean value theorem there exists  $\xi^k \in [w^k, w^{k+1}]$  such that

$$h(w^{k+1}) = h(w^k) + \langle \nabla h(w^k), w^{k+1} - w^k \rangle + \frac{1}{2} \langle \nabla^2 h(\xi^k)(w^{k+1} - w^k), w^{k+1} - w^k \rangle$$

and since the  $w^k$ 's are bounded we may find  $\alpha \geq 0$  such that

$$h(w^{k+1}) \leq h(w^k) + \langle \nabla h(w^k), w^{k+1} - w^k \rangle + \alpha \|w^{k+1} - w^k\|^2.$$

But then, since  $\|w^{k+1} - w^k\| = \lambda_k \|\tilde{w}^k - w^k\| \leq \beta \lambda_k$  for some constant  $\beta \geq 0$ , using (14) we deduce

$$[\Psi(w^{k+1}) - \Psi(w^*)] \lambda_k + [h(w^{k+1}) - h(w^k)] \leq \alpha \beta^2 \lambda_k^2.$$

In particular we get  $0 \leq h(w^{k+1}) \leq h(w^k) + \alpha \beta^2 \lambda_k^2$  which implies that  $h(w^k)$  converges. On the other hand, summing these inequalities we obtain

$$\sum_{k=0}^{\infty} [\Psi(w^{k+1}) - \Psi(w^*)] \lambda_k \leq h(w^0) + \alpha \beta^2 \sum_{k=0}^{\infty} \lambda_k^2 < \infty$$

and then the condition  $\sum \lambda_k = \infty$  implies  $\liminf \Psi(w^k) \leq \Psi(w^*)$ . Since  $w^*$  is the unique minimum of  $\Psi(\cdot)$  it follows that we may find a subsequence  $w^{k_j} \rightarrow w^*$ , and then  $\lim_k h(w^k) = \lim_j h(w^{k_j}) = 0$  from which we conclude that  $w^k \rightarrow w^*$  as claimed.  $\blacksquare$

## 7 Appendix A

Let  $\mathcal{E}$  denote the class of functions  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  which can be expressed in the form

$$\varphi(x) = \mathbb{E}(\min_{i=1\dots n} [x_i + \epsilon_i]) \quad (15)$$

for random variables  $\epsilon_i$ 's such that  $\mathbb{E}(\epsilon_i) = 0$  and

$$\mathbb{P}(\epsilon_i - \epsilon_j = \alpha) = 0 \quad \forall \alpha \in \mathbb{R}, i \neq j. \quad (16)$$

**EXAMPLE.** The logit model assumes that the  $\epsilon_i$ 's are independent Gumbel variables with parameter  $\beta$ , which gives  $\varphi(x) = -\frac{1}{\beta} \ln(e^{-\beta x_1} + \dots + e^{-\beta x_n})$ . In the probit model with normally distributed  $\epsilon_i$ 's there is no simple analytical expression for  $\varphi(x)$ .

**Proposition A.1** Every  $\varphi \in \mathcal{E}$  is concave and  $C^1$  with  $\varphi(x) \leq m(x) \triangleq \min\{x_1, \dots, x_n\}$  and

$$\frac{\partial \varphi}{\partial x_i}(x) = \mathbb{P}(x_i + \epsilon_i \leq x_j + \epsilon_j, \forall j \neq i). \quad (17)$$

*Proof.* The inequality  $\varphi(x) \leq m(x)$  follows at once by taking expectation in the inequality  $\min_{i=1\dots n} [x_i + \epsilon_i] \leq x_j + \epsilon_j$ . Let  $F(\epsilon)$  be the joint distribution of  $\epsilon = (\epsilon_1, \dots, \epsilon_n)$  so that

$$\varphi(x) = \int_{\mathbb{R}^n} m(x + \epsilon) dF(\epsilon).$$

Since  $m$  is concave the same holds for  $\varphi$ . To compute  $\frac{\partial \varphi}{\partial x_i}$  we consider the differential quotient

$$\frac{\varphi(x + te_i) - \varphi(x)}{t} = \int_{\mathbb{R}^n} q_t(\epsilon) dF(\epsilon)$$

where  $q_t(\epsilon) = [m(x + \epsilon + te_i) - m(x + \epsilon)]/t$ . Denoting

$$\begin{aligned} A &= \{\epsilon \in \mathbb{R}^n : x_i + \epsilon_i < x_j + \epsilon_j, \forall j \neq i\} \\ B &= \{\epsilon \in \mathbb{R}^n : x_i + \epsilon_i \leq x_j + \epsilon_j, \forall j \neq i\} \end{aligned}$$

it follows that  $\lim_{t \downarrow 0^+} q_t(\epsilon) = 1_A(\epsilon)$  and  $\lim_{t \uparrow 0^-} q_t(\epsilon) = 1_B(\epsilon)$ . Since the convergence is monotone we may use Lebesgue's theorem to deduce

$$\begin{aligned} D_i^+ \varphi(x) &= \int_{\mathbb{R}^n} 1_A(\epsilon) dF(\epsilon) = \mathbb{P}(A), \\ D_i^- \varphi(x) &= \int_{\mathbb{R}^n} 1_B(\epsilon) dF(\epsilon) = \mathbb{P}(B). \end{aligned} \quad (18)$$

Assumption (16) gives  $\mathbb{P}(A) = \mathbb{P}(B)$  and therefore the partial derivative  $\frac{\partial \varphi}{\partial x_i}$  exists and satisfies (17). The  $C^1$  character then follows since  $\varphi$  is concave.  $\blacksquare$

The following result of San Martín [13], provides a useful characterization of the class  $\mathcal{E}$ .

**Proposition A.2**  $\varphi \in \mathcal{E}$  iff the following properties hold:

- (a)  $\varphi$  is  $C^1$  and componentwise non-decreasing.
- (b)  $\varphi(x_1 + c, \dots, x_n + c) = \varphi(x_1, \dots, x_n) + c$ .
- (c)  $\varphi(x) \rightarrow x_i$  when  $x_j \rightarrow \infty$  for all  $j \neq i$ .

(d) for  $x_i$  fixed, the function  $\frac{\partial \varphi}{\partial x_i}(x_1, \dots, x_n)$  is a distribution on the remaining variables, i.e. it is component-wise non-decreasing with respect to  $x_j$  for  $j \neq i$  and

$$\frac{\partial \varphi}{\partial x_i}(c, \dots, c, x_i, c, \dots, c) \rightarrow \begin{cases} 1 & \text{if } c \rightarrow +\infty \\ 0 & \text{if } c \rightarrow -\infty. \end{cases}$$

*Proof.* For  $\varphi \in \mathcal{E}$  the properties (a) through (d) follow easily from (15) and Proposition A.1. To establish the converse let us consider a random vector  $\eta = (\eta_2, \dots, \eta_n)$  with distribution

$$F_\eta(x_2, \dots, x_n) = \frac{\partial \varphi}{\partial x_1}(0, x_2, \dots, x_n).$$

We begin by noting that property (b) implies

$$\begin{aligned} \varphi(x) &= x_1 - \int_a^{x_1} [1 - \frac{\partial \varphi}{\partial x_1}(y, x_2, \dots, x_n)] dy \\ &\quad + \varphi(0, x_2 - a, \dots, x_n - a) \end{aligned}$$

so that letting  $a \rightarrow -\infty$  and using (c) we get

$$\varphi(x) = x_1 - \int_{-\infty}^{x_1} [1 - \frac{\partial \varphi}{\partial x_1}(y, x_2, \dots, x_n)] dy.$$

Now, property (b) also implies

$$\begin{aligned} \frac{\partial \varphi}{\partial x_1}(y, x_2, \dots, x_n) &= \frac{\partial \varphi}{\partial x_1}(0, x_2 - y, \dots, x_n - y) \\ &= F_\eta(x_2 - y, \dots, x_n - y) \\ &= \mathbb{P}(y \leq Y) \end{aligned}$$

with  $Y = \min\{x_2 - \eta_2, \dots, x_n - \eta_n\}$ , and therefore

$$\varphi(x) = x_1 - \int_{-\infty}^{x_1} F_Y(y) dy.$$

Integration by parts allows to further work out this expression as

$$\begin{aligned} \varphi(x) &= x_1 - \int_{-\infty}^{x_1} [x_1 - y] dF_Y(y) \\ &= x_1[1 - \mathbb{P}(Y \leq x_1)] + \int_{-\infty}^{x_1} y dF_Y(y) \\ &= \int_{-\infty}^{\infty} \min\{x_1, y\} dF_Y(y) \\ &= \mathbb{E}(\min\{x_1, Y\}) \\ &= \mathbb{E}(\min\{x_1, x_2 - \eta_2, \dots, x_n - \eta_n\}) \end{aligned}$$

and then we may conclude by taking  $\epsilon_1 = 0$  and  $\epsilon_i = -\eta_i$  for  $i = 2, \dots, n$ . Notice that  $\mathbb{E}(\epsilon_i) = 0$  follows from (c) by using Lebesgue's theorem, while (16) is obtained from (18) using the fact that  $\varphi$  is differentiable. ■

REMARK. We observe that condition (d) may be weakened to  $\frac{\partial \varphi}{\partial x_1}(0, x_2, \dots, x_n)$  being a distribution on  $\mathbb{R}^{n-1}$ .

## 8 Appendix B

This short appendix provides analytic expressions for the Hessians of the functions  $\Phi(t)$  and  $\Psi(w) = \Phi(s(w))$ , which are required to implement Newton's method. We denote

$$B^d(t) = [I + J^d \frac{\partial \tau^d}{\partial t}]$$

where  $J^d = (J_{aj})_{a \in A, j \neq d}$  with  $J_{aj}^d = 1$  if  $j = j_a$  and  $J_{aj}^d = 0$  otherwise. Recall from Corollary 1 that

$$\frac{\partial \tau^d}{\partial t} = [I - \hat{P}^d(t)]^{-1} \hat{Q}^d(t)$$

**Proposition B.1** *If the functions  $\varphi_i^d(\cdot)$  are of class  $C^2$  then we have*

$$\nabla^2 \Phi(t) = \text{diag}[1/s'_a(s_a^{-1}(t_a))] - \sum_{d \in D} B^d(t)' [\sum_{i \neq d} x_i^d(t) \nabla^2 \varphi_i^d(z^d(t))] B^d(t).$$

*Proof.* Setting  $\psi^d(t) = \sum_{i \neq d} g_i^d \tau_i^d(t)$  as in Corollary 1, the formula for  $\nabla^2 \Phi(t)$  follows easily if we show that

$$\nabla^2 \psi^d(t) = (B^d)' [\sum_{i \neq d} x_i^d \nabla^2 \varphi_i^d(z^d)] B^d. \quad (19)$$

To prove the latter we exploit the equality  $[I - \hat{P}^d] \frac{\partial \tau^d}{\partial t_a} = \hat{Q}^d_{.a}$  which gives

$$\frac{\partial^2 \tau^d}{\partial t_b \partial t_a} = [I - \hat{P}^d]^{-1} \left\{ \frac{\partial \hat{Q}^d_{.a}}{\partial t_b} + \frac{\partial \hat{P}^d}{\partial t_b} \frac{\partial \tau^d}{\partial t_a} \right\}.$$

Then, using the fact that  $x^d = [I - (\hat{P}^d)']^{-1} g^d$  and the chain rule we get

$$\begin{aligned} \frac{\partial^2 \psi^d}{\partial t_b \partial t_a} &= (g^d)' [I - \hat{P}^d]^{-1} \left\{ \frac{\partial \hat{Q}^d_{.a}}{\partial t_b} + \frac{\partial \hat{P}^d}{\partial t_b} \frac{\partial \tau^d}{\partial t_a} \right\} \\ &= (x^d)' \left\{ \frac{\partial \hat{Q}^d_{.a}}{\partial t_b} + \frac{\partial \hat{P}^d}{\partial t_b} \frac{\partial \tau^d}{\partial t_a} \right\} \\ &= \sum_{i \neq d} x_i^d \left\{ \frac{\partial Q_{ia}^d}{\partial t_b} + \sum_{j \neq d} \frac{\partial P_{ij}^d}{\partial t_b} \frac{\partial \tau_j^d}{\partial t_a} \right\} \\ &= x_{i_a}^d \frac{\partial}{\partial t_b} \left( \frac{\partial \varphi_{i_a}^d}{\partial z_a^d} \right) + \sum_{i \neq d} x_i^d \sum_{c \in A_i^+} \frac{\partial}{\partial t_b} \left( \frac{\partial \varphi_i^d}{\partial z_c^d} \right) \frac{\partial \tau_{j_c}^d}{\partial t_a} \\ &= \sum_{i \neq d} x_i^d \sum_{c \in A_i^+} \frac{\partial}{\partial t_b} \left( \frac{\partial \varphi_i^d}{\partial z_c^d} \right) \left[ \delta_{ca} + \frac{\partial \tau_{j_c}^d}{\partial t_a} \right] \\ &= \sum_{i \neq d} x_i^d \sum_{c, e \in A_i^+} \frac{\partial^2 \varphi_i^d}{\partial z_e^d \partial z_c^d} \left[ \delta_{eb} + \frac{\partial \tau_{j_e}^d}{\partial t_b} \right] \left[ \delta_{ca} + \frac{\partial \tau_{j_c}^d}{\partial t_a} \right] \end{aligned}$$

where  $\delta_{ca} = 1$  if  $c = a$  and  $\delta_{ca} = 0$  otherwise. This corresponds precisely to (19) as was to be proved.  $\blacksquare$

Using the chain rule and the equality  $\nabla[\sum_{d \in D} \psi^d(t)] = \sum_{d \in D} v^d(t) = \tilde{w}(t)$  one may easily deduce

**Proposition B.2** *If the functions  $\varphi_i^d(\cdot)$  are of class  $C^2$  then we have*

$$\nabla^2\Psi(w) = M(w) + \Gamma(w)\nabla^2\Phi(s(w))\Gamma(w)$$

with

$$\begin{aligned} M(w) &= \text{diag}[s_a''(w_a)(w_a - \tilde{w}_a(s(w))) : a \in A] \\ \Gamma(w) &= \text{diag}[s_a'(w_a) : a \in A]. \end{aligned}$$

**Acknowledgements.** We sincerely thank Jaime San Martín for long discussions on the probabilistic foundation of discrete choice models as well as for the characterization of the class  $\mathcal{E}$  in Proposition A.2. We are also indebted to Marleen van Rijsbergen who helped us in implementing Newton’s method for the MTE.

## References

- [1] Ahuja R.K., Magnanti Th.L. and Orlin J.B., “Network flows”, in *Handbooks in Operations Research and Management Science, Vol. 1 Optimization*, North-Holland, Amsterdam, 211-369, 1989.
- [2] Akamatsu T., “Cyclic flows, Markov processes and stochastic traffic assignment”, *Transportation Res. Part B* 30(5), 369–386, (1996).
- [3] Alvarez F., Bolte J. and Brahic O., “Hessian Riemannian gradient flows in convex programming”, to appear in *SIAM J. Control Optim.* (2003).
- [4] Beckman M., McGuire C. and Winsten C., *Studies in Economics of Transportation*, Yale University Press, New Haven, 1956.
- [5] Daganzo C. and Sheffi Y., “On stochastic models of traffic assignment”, *Transportation Science* 11, 253–274 (1977).
- [6] Dial R.B., “A probabilistic multipath traffic assignment model which obviates path enumeration”, *Transportation Research* 5, 83–111 (1971).
- [7] Fisk C., “Some developments in equilibrium traffic assignment”, *Transportation Research* 14B, 243–255, (1980).
- [8] Florian M. and Hearn D., “Network equilibrium models and algorithms”, in *Handbooks in OR & MS*, M.O. Ball et al. Eds., Elsevier Science B.V, (1995).
- [9] Florian M. and Hearn D., “Network equilibrium and pricing”, in *Handbook of Transportation Science*, R.W. Hall Ed., Kluwer (1999).
- [10] Fukushima M., “On the dual approach to the traffic assignment problem”, *Transportation Res. Part B* 18(3), 235–245, (1984).
- [11] Larsson T., Liu Z. and Patrickson M., “A dual scheme for traffic assignment problems”, *Optimization* 42(4), 323–358, (1997).
- [12] Nguyen S., Pallottino S. and Gendreau M., “Implicit enumeration of hyperpaths in a logit model for transit networks”, *Transportation Science* 32, 54–64, (1998).
- [13] San Martín J., Personal communication, (2003).

- [14] Sheffi Y. and Powell W., “An algorithm for the equilibrium assignment with random link times”, *Networks* 12, 191–207, (1982).
- [15] Shor N.Z., *Minimization methods for non-differentiable functions*, Springer Series in Computational Mathematics, Springer-Verlag, Berlin (1985).
- [16] Trahan M., “Probabilistic assignment: an algorithm”, *Transportation Science* 8, 311–320 (1974).
- [17] Wardrop J.G., “Some theoretical aspects of road traffic research”, in *Proceedings of the Institute of Civil Engineers, Part II*, 325–378, 1952.