Optimization Formulations and Static Equilibrium in Congested Transportation Networks *

Andre de Palma† and Yurii Nesterov‡

July, 1998

Abstract

In this paper we study the concepts of equilibrium and optimum in static transportation networks with elastic and non-elastic demands. The main mathematical tool of our paper is the theory of variational inequalities. We demonstrate that this theory is useful for proving the existence theorems. It also can justify Beckmann's formulation of the equilibrium problem. The main contribution of this paper is to propose a new definition of equilibrium, the *normal equilibrium*, which exists under very general assumptions. This concept can be used, in particular, when the travel costs are discontinuous and unbounded. As examples we consider the models of signalized intersections, traffic lights and unbounded travel-time relationships. For some of those cases, the standard concepts of user and Wardrop equilibria cannot be used.

*We would like to thank FNRS Suisse, Modèle de congestion urbain for partial financial support
†THEMA, Université de Cergy-Pontoise, France
‡CORE, Université Catholique de Louvain, Belgium

1 Introduction

The description of driver's behavior in congested networks has attracted attention in economics, transportation and operations research since several decades. The study of the basic model with one origin, one destination and two roads in parallel is attributed to Pigout (1920) and was solved by Knight in 1924 (see Arnott et al. 1994). The equilibrium concept used is known in the transport literature as Wardrop First Principle (1952) which states that the travel time of all used routes are equal and less than those of the unused routes (a formal definition will be provided in Section 3.1). It is known that Wardrop First Principle (equilibrium) is the solution of a non-cooperative Cournot–Nash game. An other possible equilibrium concept is the *user equilibrium*. The transportation system reaches a user equilibrium if no infinitesimal flow on a route can swap to another route connecting the same origin-destination and experience a strictly lower cost (a similar concept was used by Smith (1979) in a pioneer paper). We should also mention the Wardrop Second Principle (optimum), which states that the average travel time is minimal. This corresponds to the first best social optimum in economic terms.

The formulation of the Wardrop equilibrium for a general network was first formulated as a mathematical programming problem by Beckmann, McGuire and Winsten (1956). Very soon the relation between Nash equilibrium and the theory of variational inequalities was established (see Kinderlehrer and Stampacchia (1980) for references). Those mathematical results were used by Dafermos and Sparrow (1969), who derived the formulation of the Wardrop equilibrium as a variational inequality problem. This approach, contrary to Beckmann et al. (1956) allows for interaction between the arc travel cost functions.[1]

The above mentioned approach is static since congestion only depends on the total number of cars crossing the arcs of the network. It has been challenged recently by dynamic models in which congestion depend on the time of the day (see the seminal paper by Vickrey, 1969). However, most (if not all) computational tools in transportation practice up to now are still static, so that static analysis still remains a focus point in transportation research. Later on, it will remain at least a useful benchmark. Finally, many other applications of static network equilibrium have been developed outside the transportation field (in spatial economics and in finance among others, see Nagurney (1993) for details).

As we have seen, the classical theory for static transportation models is rather old. However, during last years we can see a revival of this field. The reason is that in practice several new tools appear to be useful in controlling congestion (tolls, traffic lights, etc.). These tools (Emme 2, Polydrome, Davis, etc.) very often cannot be treated by the existing models which deal with

[1] We refer the reader to Sheffi 1985 for a clear synthesis of the mathematical programming formulation. The variational inequality approach is extensively studied in the book by Nagurney (1993).

smooth functions defined for the all possible values of the flows. Thus, there is a need to generalize the existing theory to take into account these new requirements. Our paper is one of the attempts to feel these gaps.

In the proposed approach we would like to introduce a unifying framework which allows to incorporate the following features:

- We allow travel cost functions to be discontinuous. This feature is important when dealing with congestion pricing. First because a socially optimal congestion pricing which continuously depends on road usage cannot be implemented. Second because, as we shall show in Section 7.2, the optimal congestion pricing may itself be discontinuous. Discontinuous travel cost functions were studied by Bernstein (1994). However, his analysis is based on rather specific assumptions (pairwise monotonicity) which breaks down the possibility to use the standard mathematical programming theory and numerical methods.
- We allow travel costs to be unbounded on a finite feasible interval. More precisely, we can consider the travel cost functions, which go to infinity when the load reaches the upper bound of the interval. Several traffic flow functions (as the Greenshield relation) used by theorists or practitioners assume that speed is zero if road occupancy reaches a critical level. We can use this type of travel functions forbidden in existing models.

We describe the congestion technology by monotone operators, which in general can be multivalued. This structure allows to treat many features of the real networks including junctions and other interactions.

In order to incorporate the above features in the analysis, we introduce a new equilibrium concept that we call *normal* equilibrium. We show under which conditions normal equilibrium reduces to Wardrop equilibrium and to user equilibrium.

The paper is organized as follows. In Section 2 we introduce the main notation and the general setting of the formulations. Section 3 includes definitions and the main theorems concerned with variational inequalities which we use later on. We study three definitions of the equilibrium in the model with constant demand: Wardrop equilibrium, user equilibrium and normal equilibrium, under very mild assumption on the travel cost operator: it can be discontinuous, unbounded and dependent on the flows on other arcs. We demonstrate first that the normal equilibrium very often exists even when the Wardrop or the user equilibrium does not exist (the opposite situation is impossible). Normal equilibrium is a useful concept since it (and only it) can be computed by standard numerical methods.

The structure of Section 4 is similar to that of the previous section; it is devoted to elastic demand. In Section 5 we study the possibilities to find an equilibrium in a transportation network as a solution of a minimization problem. We present the corresponding minimization problems for finding the normal equilibrium with constant and elastic demand. We propose an intuitive interpretation of Beckmann's formulation and establish a formal relation with the theory of potential operators.

Section 6 briefly discusses the system (or social) optimum problem. Finally, in Section 7 we treat three examples of application of the developed theory. In Section 7.1 we consider a model of intersection with traffic lights. In this model, the travel time functions on the incoming arcs depend on the flows on other arcs. Section 7.2 is devoted to the analysis of congestion toll. We demonstrate that the social optimum can be reached by a system of tolls and that such tolls can be discontinuous. We also argue that in the models we propose the social optimum cannot be reached by traffic light regulation. In Section 7.3 we discuss a model of traffic congestion with unbounded travel time. This feature reflects the bounded capacities of the real roads. Section 8 concludes the paper.

2 Description of the model

In this paper we deal with a general transportation network $\mathcal{G}$ with continuous number of users. This network consists of a set of nodes and a set of directed arcs. We will use the following notation:

$\mathcal{N}$ – the set of nodes;

$\mathcal{A}$ – the set of directed arcs, $A :=\mid \mathcal{A} \mid$; (i.e. A is the cardinality of $\mathcal{A}$);

$\mathcal{P} = \{(n_o, n_d),\ n_o \in \mathcal{N}, n_d \in \mathcal{N}\}$ – the set of origin-destination pairs; $P :=\mid \mathcal{P} \mid$.

We assume that any O/D pair is connected;

$\mathcal{H}_p$ – the set of paths for the O/D pair number p; $M_p :=\mid \mathcal{H}_p \mid > 0$;

M - the total number of paths in the network; $M = \sum\limits_{p=1}^{P} M_p$;

$\mathcal{R}^p_k$ – the path number k for the O/D pair number p, $k = 1 \dots M_p$, $p = 1 \dots P$;

$\delta^p_{a,k} = \begin{cases} 1, & \text{if arc } a \in \mathcal{R}^p_k, \\ 0, & \text{otherwise.} \end{cases}$, $a \in \mathcal{A}$, $k \in \mathcal{H}_p$, $p \in \mathcal{P}$;

x_a – the flow on the arc a, for $a \in \mathcal{A}$;

x – the arc flow vector; $x = (x_1 \dots x_a \dots x_A)^T$;

$T_a(x)$ – the travel cost function of the arc a, $a \in \mathcal{A}$, for the link flow vector x;

$\mathcal{D}_a$ – an open domain of $T_a(x)$;

$T(x)$ – the travel cost operator; $T(x) = (T_1(x) \dots T_a(x) \dots T_A(x))^T$;

$\mathcal{D}$ – an open domain of $T(x)$;

f^p_k – the flow in the path k of the O/D pair number p, $k = 1 \dots M_p$, $p = 1 \dots P$;

F_p – the flow vector of the O/D pair number p; $F_p = (f^p_1 \dots f^p_k \dots f^p_{M_p})^T$;

F – the whole (column) flow vector in $\mathcal{G}$; $F = (F_1 \dots F_p \dots F_P)$;

$\mathcal{F}$ – the set of feasible flows;

c^p_k – the cost of path k of the O/D pair number p, $k = 1 \dots M_p$, $p = 1 \dots P$;

C_p – the cost vector of the O/D pair number p; $C_p = (c^p_1 \dots c^p_k \dots c^p_{M_p})^T$;

C – the whole (column) cost vector in $\mathcal{G}$; $C = (C_1 \dots C_p \dots C_P)$;
u_p^* - the equilibrium cost for the O/D pair number p, $p = 1 \dots P$;
u^* - the equilibrium cost vector; $u^* = (u_1^* \dots u_p^* \dots u_P^*)^T$;
q^p – the demand of the O/D pair number p, $p = 1 \dots P$;
q – the demand vector; $q = (q^1 \dots q^p \dots q^P)^T$;
$\Pi^p(q)$ – the equilibrium cost function for the O/D pair p given the demand vector
q, $p = 1 \dots P$;
$\Pi(q)$ – the equilibrium cost operator; $\Pi(q) = (\Pi^1(q) \dots \Pi^p(q) \dots \Pi^P(q))^T$.

Moreover, we will use the following definitions. For two vectors $x, y \in E$, where E is an n-dimensional real vector space we denote the *inner product* as:

$$\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i,$$

where x_i, y_i are the ith components of these vectors. We also use notation B^T for the transposed matrix of B. The following trivial property will be useful

$$\langle B^T x, y \rangle = \langle x, By \rangle.$$

The notation $x \cdot y$ for $x, y \in E$, is used for the vector from E with components $x_i y_i$, $i = 1 \dots n$. Finally, the inequality $x \geq y$, where $x, y \in E$, means that $x_i \geq y_i$, $i = 1 \dots n$.

For a differentiable function $f(x) : E \to R$, we denote

$$f'(x) = \left(\frac{\partial f(x)}{\partial x_1} \dots \frac{\partial f(x)}{\partial x_n} \right)^T.$$

Let $F(x) : E \to E_1$ be a vector function, where E_1 is an m-dimensional real vector space. We denote the Jacobian of $F(x)$ by

$$F'(x) = \begin{pmatrix} \frac{\partial F_1(x)}{\partial x_1} & \dots & \frac{\partial F_1(x)}{\partial x_n} \\ \dots & \dots & \dots \\ \frac{\partial F_m(x)}{\partial x_1} & \dots & \frac{\partial F_m(x)}{\partial x_n} \end{pmatrix}.$$

Using the notation we can write the flow conservation in the network as follows:

$$q^p = \sum_{k=1}^{M_p} f_k^p, \quad p = 1 \dots P,$$

or, using a matrix notation, as

$$q = BF, \tag{1}$$

where B is the corresponding $(P \times M)$-matrix with 0-1 entries, which is organized as follows: pth row of this matrix has all zero entries except that ones, which corresponds to the position of the sub-vector F_p in the flow vector F (the latter entries are equal to one). We require the non-negativity constraint

$$F \geq 0. \tag{2}$$

Similarly, the expressions for arc flows can be written as

$$x_a = \sum_{p=1}^{P} \sum_{k=1}^{M_p} f_k^p \delta_{a,k}^p, \quad a \in \mathcal{A},$$

or, in the matrix notation

$$x = \Psi F, \tag{3}$$

where Ψ is the corresponding $(A \times M)$-matrix:

$$\Psi_{a,j(k,p)} = \delta_{a,k}^p,$$

where $j(k,p)$ is the position of the element f_k^p in the flow vector F. The picture of this matrix looks as follows:

$$\Psi = \begin{pmatrix} \delta_{1,1}^1 & \dots & \delta_{1,M_P}^P \\ \dots & \delta_{a,k}^p & \dots \\ \delta_{A,1}^1 & \dots & \delta_{A,M_P}^P \end{pmatrix}.$$

Thus, the set of *feasible* flows is defined as

$$\mathcal{F} = \{F : \ F \geq 0, \ BF = q, \ \Psi F \in \mathcal{D}\}.$$

Finally, for the cost of the path k of O/D pair number p we have

$$c_k^p = \sum_{a \in \mathcal{A}} T_a(x) \delta_{a,k}^p, \quad x \in \mathcal{D}, \tag{4}$$

or, in the matrix notation

$$C = \Psi^T T(x), \quad x \in \mathcal{D}. \tag{5}$$

In what follows, we will use notation $x(F)$ for the arc flow x given by (3). Similarly, the notation $C(F)$ will be used for the cost vector C, which is defined by (5) using $x = x(F)$. The same notation will be used for the components of these vectors, $x_a(F)$ and $c_k^p(F)$.

The specific form of the equations (1)-(5) is the basis for our analysis.

3 The equilibrium concepts for constant demand

3.1 Wardrop and user equilibrium

There exist two main equilibrium concepts. The first one is the traditional *Wardrop Equilibrium* (see Wardrop (1952), [11]):

Definition 1 *The flow pattern F^*, which is feasible for (1)-(5), is called the* Wardrop equilibrium *if for any O/D pair p the costs for all used paths are the same and there are no unused paths with strictly smaller cost.*

Mathematically this concept can be written in the following way: there exists an equilibrium cost vector u^* such that for all path costs c_k^p (defined by (3) and (4) with some $F = F^*$) we have

$$c_k^p \geq u_p^*, \quad k = 1 \dots M_p, \quad p = 1 \dots P,$$

and

$$f_k^p(c_k^p - u_p^*) = 0, \quad k = 1 \dots M_p, \quad p = 1 \dots P.$$

Using our matrix notation we can rewrite these conditions as follows:

$$C^* \geq B^T u^*, \quad F^* \cdot (C^* - B^T u^*) = 0, \tag{6}$$

where from (5), $C^* = \Psi^T T(x^*)$, and from (3), $x^* = \Psi F^*$.

The second important concept is the *user equilibrium* (see, e.g. [3]):

Definition 2 *The flow pattern $F^* \in \mathcal{F}$ is called the* user equilibrium *if for any O/D pair p and any path $k_1 \in \mathcal{H}_p$ with $f_{k_1}^p(F^*) > 0$*

$$c_{k_1}^p(F^*) \leq \liminf_{\epsilon \to 0^+} c_{k_2}^p(F^* - \epsilon 1_{k_1}^p + \epsilon 1_{k_2}^p), \quad \forall k_2 \in \mathcal{H}_p, \tag{7}$$

where 1_k^p is the unit coordinate vector, which corresponds to the position of the component f_k^p in a flow vector F.

This definition implies that for each O/D pair p, no arbitrary small pack of drivers on any path $\mathcal{R}_k^p$ can strictly decrease their cost by switching to any other path $\mathcal{R}_{k'}^p$, $k' \neq k$. Note that for continuous operator $T(x)$ the notion of Wardrop and user equilibrium clearly coincides.

In the general case, the relation between Wardrop and user equilibrium is rather complicated: it is not difficult to find examples for which a Wardrop equilibrium exists and a user equilibrium does not exist and vice versa. This is illustrated by the example below.

Example 1 *Let the network $\mathcal{G}$ consists of one O/D pair, which is connected by two arcs. Let the travel cost function for the first arc is*

$$T_1(x_1) = x_1.$$

Consider for the second arc the travel cost function

$$T_2(x_2) = \begin{cases} \frac{5}{4}x_2, & \text{if } x_2 < \frac{1}{2}, \\ \frac{1}{2}, & \text{if } x_2 = \frac{1}{2}, \\ \frac{5}{4}x_2 - \frac{1}{4}, & \text{if } x_2 > \frac{1}{2}. \end{cases}$$

Assume that the demand q^1 for the network is equal to 1. Then

$$f_1^1 = x_1 = \frac{1}{2}, \quad f_2^1 = x_2 = \frac{1}{2} \tag{8}$$

is a Wardrop equilibrium since in this case

$$T_1(x_1) = T_2(x_2) = \frac{1}{2}.$$

However, it is not a user equilibrium since any driver using the first arc can strictly decrease his cost by shifting on the second arc. Mathematically this means that in (7) we have

$$\lim_{\epsilon \to 0^+} c_2^1(x_2 + \epsilon) = \tfrac{3}{8} < \tfrac{1}{2} = c_1^1(x_1).$$

Example 2 *Consider now* $T_1(x_1) = x_1$ *and*

$$T_2(x_2) = \begin{cases} \frac{1}{2}x_2, & \text{if } x_2 \le \frac{1}{2}, \\ \frac{1}{2}x_2 + \frac{1}{2}, & \text{if } x_2 > \frac{1}{2}. \end{cases}$$

Then the point defined by (8) is not a Wardrop equilibrium (it is easy to see that there is no Wardrop equilibrium in this example). However, this point is a user equilibrium since a driver from the first arc has no incentive to switch on the other arc (this would strictly increase his cost).

Example 3 *Finally, if* $T_1(x_1) = x_1$ *and*

$$T_2(x_2) = \begin{cases} \frac{1}{2}x_2, & \text{if } x_2 < \frac{1}{2}, \\ \frac{5}{8}, & \text{if } x_2 = \frac{1}{2}, \\ \frac{1}{2}x_2 + \frac{1}{2}, & \text{if } x_2 > \frac{1}{2}, \end{cases}$$

then neither a Wardrop nor a user equilibrium does exist.

The above examples suggest that we need additional assumptions to guarantee the existence of an equilibrium.

Assumption 1 *a). The travel cost operator* $T(x)$, $x \ge 0$, *is single-valued and monotone:*

$$\langle T(x) - T(y), x - y \rangle \ge 0 \quad \forall x, y \in \mathcal{D}.$$

b). The operator $T(x)$ *is complete: For any* $y \in \mathcal{D}$

$$\langle T(x), x - y \rangle \to +\infty \quad \textit{as} \quad x \to \partial\mathcal{D}.$$

(See Appendix for some theoretical results on such operators.)

It is easy to check that this assumption implies the monotonicity and completeness of the operator $C(F)$ since in view of (3) and (5)

$$C(F) = \Psi^T T(\Psi F), \quad \mathrm{dom}\, C = \{F : \ \Psi F \in \mathcal{D}\}.$$

As we will see later, this assumption also facilitates the numerical computation of the equilibrium problem. Note that Examples 2 and 3 deal with monotone operator $T(x) = (T_1(x_1), T_2(x_2))^T$ and the operator $T(x)$ in Example 1 is not monotone since the function $T_2(x_2)$ is not monotonically increasing at $x_2 = \frac{1}{2}$. All these operators are not continuous.

Example 4 *Assume that the operator $T(x) = (T_1(x) \dots T_A(x))^T)$ is differentiable on $\mathcal{D}$. Then $T(x)$ is monotone if and only if for any $x \in \mathcal{D}$ the Jacobian $T'(x)$ is a positive semidefinite matrix:*

$$\langle T'(x)s, s\rangle \geq 0, \quad \forall s \in E. \tag{9}$$

If for any $s \neq 0$ this inequality is strict, then $T(x)$ is strictly monotone. In 2-dimensional case condition (9) is equivalent to

$$\frac{\partial T_1(x_1,x_2)}{\partial x_1} \geq 0, \qquad \frac{\partial T_2(x_1,x_2)}{\partial x_2} \geq 0,$$

$$4\frac{\partial T_1(x_1,x_2)}{\partial x_1}\frac{\partial T_2(x_1,x_2)}{\partial x_2} \quad \geq \quad \left(\frac{\partial T_1(x_1,x_2)}{\partial x_2} + \frac{\partial T_2(x_1,x_2)}{\partial x_1}\right)^2.$$

Theorem 1 *Let the travel cost operator $T(x)$ satisfies Assumption 1. Then*

(i) the set of Wardrop equilibria in the network $\mathcal{G}$ coincides with the set of strong solutions to the following variational inequality:

$$\langle C(F), F - F^*\rangle \geq 0, \quad \forall F \in \mathcal{F}. \tag{10}$$

(ii) if the operator $T(x)$ is continuous, then a Wardrop equilibrium in $\mathcal{G}$ exists.

(iii) if the travel cost operator is strictly monotone and if there exists a Wardrop equilibrium, then there is a unique equilibrium in arc flow.

Proof:

(i) Let $F^* \in \mathcal{F}$ be a Wardrop equilibrium in $\mathcal{G}$. Then for any feasible F and $C^* = C(F^*)$ we have

$$\begin{aligned}
\langle C^*, F\rangle \geq \quad & \langle B^T u^*, F\rangle && \text{(by (6))} \\
= \quad & \langle u^*, BF\rangle && \text{(property of the inner product)} \\
= \quad & \langle u^*, BF^*\rangle && \text{(by (1))} \\
= \quad & \langle B^T u^*, F^*\rangle && \text{(property of the inner product)} \\
= \quad & \langle C^*, F^*\rangle && \text{(by (6)).}
\end{aligned}$$

Vice versa, let $(F^*, C(F^*))$ be a strong solution to (10). Then, in view of Lemma 6, F^* is the solution of the following Linear Programming problem:

$$\begin{array}{cl}
\text{Minimize:} & \langle C^*, F\rangle \\
F & \\
\text{Subject to:} & \\
 & BF = q, \\
 & F \geq 0,
\end{array}$$

where $C^* = C(F^*)$. Since the feasible set of this problem is nonempty and bounded, the solution of this problem is a saddle point of the Lagrangian:

$$\mathcal{L}(F, u, v) = \langle C^*, F\rangle + \langle u, q - BF\rangle - \langle v, F\rangle,$$

where the elements of the vector u are the dual variables, and the elements of the vector v, $v \geq 0$, are the slack variables. Let (F^*, u^*, v^*) be a saddle point of the Lagrangian. Let us write the Kuhn-Tucker conditions for the linear problem. By differentiating $\mathcal{L}(F, u, v)$ with respect to F at the saddle point we obtain

$$C^* = B^T u^* + v^*.$$

The complementarity conditions for this problem can be written as $v^* * F^* = 0$. Since $C^* = C(F^*)$, we conclude that F^* is the Wardrop equilibrium (see definition (6)).

(ii) The statement of this part follows from the previous item and Theorem 12(2).

Part (iii) also follows from Lemma 4. □

Part (iii) of the above theorem states that the equilibrium arc flow is unique given that the travel cost operator $T(x)$ is strictly monotone. Note that in general this does not imply that the $C(F) = \Psi^T T(\Psi F)$ is a strictly monotone operator. Therefore the equilibrium flow pattern F^* may be not unique. This can be seen from the following example.

Example 5 *Consider a network $\mathcal{G}$ which has three nodes n_i, $i = 1 \dots 3$. Nodes n_1 and n_2 are connected by two parallel arcs a_1 and a_2. Nodes n_2 and n_3 are also connected by parallel arcs a_3 and a_4. The single O/D pair in this network is (n_1, n_3). It is clear that there are four possible paths in $\mathcal{G}$:*

$$\mathcal{R}_1 = (a_1, a_3), \quad \mathcal{R}_2 = (a_1, a_4), \quad \mathcal{R}_3 = (a_2, a_3), \quad \mathcal{R}_4 = (a_2, a_4).$$

Denote by F_i the feasible flow for the path $\mathcal{R}_i$, $i = 1 \dots 4$. Let $x^* = (x_1^* \dots x_4^*)$ be an equilibrium arc flow in this network such that $x_j^* = \alpha > 0$, $j = 1 \dots 4$, (we have a solution of this type if $T_1(x) = T_2(x)$ and $T_3(x) = T_4(x)$). Note that this arc flow can be obtained either from the O/D-flows

$$F_1 = \alpha, \quad F_2 = 0, \quad F_3 = \alpha, \quad F_4 = 0,$$

or from the O/D-flows

$$F_1 = 0, \quad F_2 = \alpha, \quad F_3 = 0, \quad F_4 = \alpha.$$

It is easy to see that both variants are the equilibrium O/D-flows for our model.

As we have seen from our example, a Wardrop and a user equilibrium sometimes do not exist even if the cost functions are monotone but not continuous everywhere. In the next section we will introduce another definition of an equilibrium, which existence follows straightforwardly from the monotonicity and completeness of $T(x)$.

3.2 Normal equilibrium

We start from the definition of *normal equilibrium.*

Definition 3 *A flow $F^* \in \mathcal{F}$ is called a* normal equilibrium *of the network $\mathcal{G}$ if for any $F \in \mathcal{F}$ we have*

$$\langle C(F), F - F^* \rangle \geq 0. \tag{11}$$

Theorem 2 *Let the travel cost operator satisfies Assumption 1 and $\mathcal{F} \neq \emptyset$. Then in the network $\mathcal{G}$ there exists a normal equilibrium. If the operator $T(x)$ is strictly monotone, then the normal equilibrium in arc flow is unique.*

Proof:
Note that the feasible set defined by $Q = \{F: \ BF = q, \ \geq 0\}$ is convex, closed and boubed. Therefore, in view of Theorem 12 the normal equilibrium does exist; that is a solution to variational inequality (11).

In arc flows the variational inequality (11) can be written as follows:

$$\langle T(x), x - x^* \rangle \geq 0 \quad x \in \hat{Q} \bigcap \mathcal{D}, \tag{12}$$

where $\hat{Q} = \{x: \ \exists F \geq 0, \ BF = q, \ \Psi F = x\}$ is a closed bounded convex set. If $T(x)$ is strictly monotone, the solution to (12) is unique in view of Lemma 4. $\square$

Let us discuss some relations between the normal, the Wardrop and the user equilibria. It is clear that if the travel cost operator $T(x)$ is continuous, then all the concepts coincide. Thus, our analysis is not trivial only for discontinuous $T(x)$.

Theorem 3 *Let the travel cost operator $T(x)$ be complete and monotone. Then the Wardrop equilibrium in the network $\mathcal{G}$ exists if and only if F^* is a normal equilibrium and the pair $(F^*, C(F^*))$ is a strong solution to the variational inequality (11).*

Proof:
The proof is evident in view of Theorem 1 (i). $\square$

Thus, if a Wardrop equilibrium in the network $\mathcal{G}$ exists, then it is a normal equilibrium. The relation between the normal and the user equilibrium can be made more precise by using additional assumptions on the semi-continuity of the operator $T(x)$.

Theorem 4 *(i) Assume that the travel cost operator $T(x)$ is upper semicontinuous, i.e. for any feasible $\bar{x}$*

$$\liminf_{x \to \bar{x}} T(x) \geq T(\bar{x}).$$

Then any normal equilibrium in the network $\mathcal{G}$ is a user equilibrium.

(ii) Assume that the travel cost operator $T(x)$ is lower semicontinuous, i.e. for any feasible $\bar{x}$

$$\limsup_{x \to \bar{x}} T(x) \le T(\bar{x}).$$

Then any user equilibrium in the network $\mathcal{G}$ is a Wardrop equilibrium and a normal equilibrium.

Proof:

(i) Let F^* be a normal equilibrium. Then in view of Definition 3 we have

$$\langle C(F), F - F^* \rangle \ge 0, \quad BF = q, \ F \ge 0, \ \Psi F \in \mathcal{D}. \tag{13}$$

Let us fix an O/D pair number p and let $f_k^p(F^*) > 0$ for some k. Consider

$$F(\epsilon) = F^* + \epsilon 1_l^p - \epsilon 1_k^p,$$

where l is a number of another path for O/D pair p and ϵ is small enough (that means that $F(\epsilon) \ge 0$). Then in view of (13) we have

$$0 \le \langle C(F(\epsilon)), F(\epsilon) - F^* \rangle = \epsilon(c_l^p(F(\epsilon)) - c_k^p(F(\epsilon))).$$

Therefore

$$\liminf_{\epsilon \to 0^+} c_l^p(F(\epsilon)) \ge \liminf_{\epsilon \to 0^+} c_k^p(F(\epsilon)) \ge c_k^p(F^*).$$

(ii) Since $T(x)$ is lower semicontinuous then $C(F)$ is also lower semicontinuous. Then any user equilibrium F^* is the Wardrop equilibrium (argumentation of [3]):

$$c_k^p(F^*) \le \liminf_{\epsilon \to 0^+} c_l^p(F^* + \epsilon 1_l^p - \epsilon 1_k^p) \le \limsup_{\epsilon \to 0^+} c_l^p(F^* + \epsilon 1_l^p - \epsilon 1_k^p) \le c_l^p(F^*).$$

The last statement of this part is a direct consequence of Theorem 3.

□

The above statements demonstrate that the normal equilibrium very often coincides with the user and the Wardrop equilibria. Since the traditional equilibria sometimes do no exist, the normal equilibrium seem to be a reasonable goal for numerical methods.

4 Elastic demand

4.1 Equilibrium concepts

Very often the demand vector q in the network $\mathcal{G}$ shall be considered as a function of the travel cost: we expect the demand q^p for O/D-pair p to be a decreasing function of the minimal travel costs in the networks. That is, we allow drivers to modify their destinations as a function of travel conditions. In this case the equality constraints for the feasible set of link flows are exactly the same as (1) - (5):

$$BF = q, \quad F \geq 0, \quad x = \Psi F, \quad C = \Psi^T T(x),$$

but now q is not fixed; it is a variable in the drivers' decision process.

The qualitative description of this model can be as follows. For each O/D-pair number p we consider two types of drivers: the *active drivers* (q^p is the number of such drivers in the network) and the *potential drivers*. Active drivers are traveling in the network and use O/D pair p and potential drivers prefer to select another destination, or use another mode (e.g. public transport). We assume for simplicity that the total number of drivers for an O/D-pair number p is bounded by a constant $\bar{q}^p$. Each driver is free in changing his/her status. We assume that the decision to travel or to stay at home is based on his/her *reservation cost*: active drivers are those with a reservation cost larger than or equal to $\Pi^p(q)$. If the minimal travel cost for O/D-pair number p is greater than $\Pi^p(q)$, then there are active drivers willing to become potential drivers. If it is smaller, then some potential drivers will become active (provided that the number of the active drivers is less than $\bar{q}^p$). It is natural to assume that $\Pi^p(q)$ is decreasing in q^p (the attraction of the destination decreases if the number of active drivers increase). Now let us give the definition of the *generalized Wardrop equilibrium* in this model.

Definition 4 *We call the flow-demand pattern (F^*, q_*) a generalized Wardrop equilibrium (with elastic demands) if the flow F^* is the Wardrop equilibrium in the network $\mathcal{G}$ with demand vector q_* and for any O/D-pair number p the minimum cost u_p^* satisfies:*
a) $\Pi^p(q_*) \leq u_p^*$ if $q_*^p = 0$;
b) $\Pi^p(q_*) = u_p^*$ if $0 < q_*^p < \bar{q}^p$;
c) $\Pi^p(q_*) \geq u_p^*$ if $q_*^p = \bar{q}^p$.

Mathematically the problem of finding a Wardrop equilibrium with elastic demand can be written as follows:

Find an equilibrium flow pattern F^*, the equilibrium cost vector u^* and the equilibrium demand vector q_* such that:

$$BF^* = q_*, \quad F^* \geq 0, \quad x^* = \Psi F^* \in \mathcal{D}, \quad C^* = \Psi^T T(x^*), \quad 0 \leq q_* \leq \bar{q},$$

and

$$C^* \geq B^T u^*, \tag{14}$$

$$F^* \cdot (C^* - B^T u^*) = 0, \tag{15}$$

$$\text{if } u_p^* > \Pi^p(q_*), \text{ then } q_*^p = 0, \tag{16}$$

$$\text{if } u_p^* < \Pi^p(q_*), \text{ then } q_*^p = \bar{q}^p, \tag{17}$$

Note that from (16), (17) we have

$$\text{if } 0 < q_*^p < \bar{q}^p \text{ then } u_p^* = \Pi^p(q_*). \tag{18}$$

Note also, that for any feasible q we have

$$\Pi^p(q_*)(q^p - q^p_*) \le u^*_p(q^p - q^p_*). \tag{19}$$

Indeed, if $q^p_* = 0$ then it is true in view of a). If $0 < q^p_* < \bar{q}^p$, this is true in view of b). Now, let $q^p_* = \bar{q}^p$; then (19) is true in view of c).

Definition 5 *We call the flow-demand pattern (F^*, q_*) a generalized user equilibrium (with elastic demands) if*

(i) the flow F^* is the user equilibrium in the network $\mathcal{G}$ with demand vector q_*;

(ii) for any O/D-pair number p such that $q^p_* < \bar{q}^p$ and any $k \in \mathcal{H}_p$

$$\limsup_{\epsilon \to 0^+} [\Pi^p(q_* + \epsilon 1_p) - c^p_k(F^* + \epsilon 1^p_k)] \le 0; \tag{20}$$

(iii) for any O/D-pair number p such that $q^p_* > 0$ and for any path $k \in \mathcal{H}_p$ with $f^p_k > 0$, we have

$$\liminf_{\epsilon \to 0^+} [\Pi^p(q_* - \epsilon 1_p) - c^p_k(F^* - \epsilon 1^p_k)] \ge 0. \tag{21}$$

The idea of this definition is rather natural. First, the flow pattern at the generalized user equilibrium shall be a user equilibrium for the constant demand q_*. Second, there is no potential user, who wants to become active (condition (20)). Finally, there is no active user, who wants to become potential (condition (21)).

Let us consider now the concept of the normal equilibrium.

Definition 6 *We call the flow-demand pattern (F^*, q_*) a generalized normal equilibrium (with elastic demands) if it is a solution of the following variational inequality:*

$$\langle C(F), F - F^* \rangle \ge \langle \Pi(q), q - q_* \rangle \tag{22}$$

subject to constraints

$$\Psi F \in \mathcal{D}, \quad BF = q, \quad F \ge 0, \quad 0 \le q \le \bar{q}.$$

In what follows we study the above concepts of the equilibrium under the following assumption.

Assumption 2 *The travel time operator $T(x)$ satisfies Assumption 1 and the reservation cost operator $\Pi(q)$, defined for all q, is single-valued and anti-monotone.*

Note that under this assumption, the relations (20), (21) in the definition of the user equilibrium are equivalent, in view of Lemma 2, to the following:

$$\lim_{\epsilon \to 0^+} \Pi^p(q_* + \epsilon 1_p) \le \lim_{\epsilon \to 0^+} c^p_k(F^* + \epsilon 1^p_k), \tag{23}$$

$$\lim_{\epsilon \to 0^+} \Pi^p(q_* - \epsilon 1_p) \ge \lim_{\epsilon \to 0^+} c^p_k(F^* - \epsilon 1^p_k)]. \tag{24}$$

Clearly, when the demand vector q is constant, the generalized normal equilibrium coincides with the Wardrop, user and normal equilibrium, respectively.

4.2 Existence results

Let us start from a result on the existence of the generalized Wardrop equilibrium.

Theorem 5 *Let the travel cost operator $T(x)$ be complete and monotone and the reservation cost operator $\Pi(q)$ be antimonotone. Then*

(i) the set of the generalized Wardrop equilibria in the network $\mathcal{G}$ coincides with the set of the strong solutions $(F^*, C(F^*), q_*, -\Pi(q_*))$ of the following variational inequality:

$$\begin{aligned} &\langle C(F), F - F^* \rangle + \langle -\Pi(q), q - q_* \rangle \geq 0, \\ &\forall F : \Psi F \in \mathcal{D},\ BF = q,\ F \geq 0,\ 0 \leq q \leq \bar{q}. \end{aligned} \tag{25}$$

(ii) if the operators $T(x)$ and $\Pi(q)$ are continuous, then a generalized Wardrop equilibrium in $\mathcal{G}$ exists.

(iii) if the travel cost operator is strictly monotone, the reservation cost operator is strictly anti-monotone and if there exists a generalized Wardrop equilibrium, then this equilibrium in arc flow is unique.

Proof:

(i) Let (F^*, q_*, u^*) be the generalized Wardrop equilibrium in $\mathcal{G}$. Then for any feasible F, q and $C^* = C(F^*)$, $\Pi^* = \Pi(q_*)$, we have

$$\begin{aligned} \langle C^*, F \rangle - \langle \Pi^*, q \rangle \geq\ & \langle B^T u^*, F \rangle - \langle u^*, q - q_* \rangle - \langle \Pi^*, q_* \rangle && \text{(by (14), (19))} \\ =\ & \langle u^*, BF \rangle - \langle u^*, q - q_* \rangle - \langle \Pi^*, q_* \rangle && \text{(inner product)} \\ =\ & \langle u^*, q_* \rangle - \langle \Pi^*, q_* \rangle && (BF = q) \\ =\ & \langle C^*, F^* \rangle - \langle \Pi^*, q_* \rangle && \text{(by (15))} \end{aligned}$$

Vice versa, let $(F^*, C(F^*), q_*, -\Pi(q_*))$ be a strong solution to the variatinal inequality (25). Then in view of Lemma 6 (F^*, q_*) is the solution of the following Linear Programming problem:

$$\begin{array}{cl} \text{Minimize:} & \langle C^*, F \rangle - \langle \Pi^*, q \rangle \\ (F, q) & \\ \text{Subject to:} & \\ & BF = q, \\ & F \geq 0, \\ & 0 \leq q \leq \bar{q}. \end{array}$$

Since the feasible set of this problem is nonempty and bounded, the solution of this problem is a saddle point of the Lagrangian:

$$\mathcal{L}(F, q, u, v, w, y) = \langle C^*, F \rangle - \langle \Pi^*, q \rangle + \langle u, q - BF \rangle - \langle v, F \rangle - \langle w, q \rangle + \langle y, q - \bar{q} \rangle,$$

where the elements of the vector u are the dual variables, and the elements of the vectors v, $v \geq 0$, w, $w \geq 0$ and y, $y \geq 0$, are the slack variables. Let $(F^*, q_*, u^*, v^*, w^*, y^*)$ be a saddle point of the Lagrangian. Let us write the Kuhn-Tucker conditions for the linear problem. By differentiating $\mathcal{L}(F, q, u, v, w, y)$ with respect to F and q at the saddle point we obtain

$$C^* = B^T u^* + v^*, \quad u^* = \Pi^* + w^* - y^*.$$

And the strict complementarity conditions for this problem can be written as

$$v^* * F^* = 0, \quad w^* * q_* = 0, \quad y^* * (q_* - \bar{q}) = 0.$$

We conclude that (F^*, q_*, u^*) is the generalized Wardrop equilibrium (see definitions (14)-(17)).

(ii) The statement of this part is a direct consequence of Theorem 12 and Lemma 3.

Part (iii) also follows from Theorem 12. □

Let us present now an existence result for the generalized normal equilibrium.

Theorem 6 *A generalized normal equilibrium exists in the network $\mathcal{G}$ if the travel cost operator $T(x)$ is complete and monotone and the reservation cost operator $\Pi(q)$ is anti-monotone. If $T(x)$ is strictly monotone and $\Pi(q)$ is strictly anti-monotone, then the generalized normal equilibrium is unique in arc flow.*

Proof:
Note that the feasible set defined by

$$BF = q, \quad F \geq 0, \quad 0 \leq q \leq \bar{q},$$

is clearly bounded. Therefore, in view of Theorem 12 the normal equilibrium does exist. The uniqueness of arc flows can be proved in the same way as in Theorem 2. □

Let us discuss some relations between the generalized normal equilibrium and the generalized Wardrop and user equilibria. Obviously, if operators $T(x)$ and $\Pi(q)$ are continuous, all concepts coincide.

Theorem 7 *Let the travel cost operator $T(x)$ be complete and monotone and the reservation price operator $\Pi(q)$ be anti-monotone. Then the generalized Wardrop equilibrium in the network $\mathcal{G}$ exists if and only if (F^*, q_*) is the generalized normal equilibrium and the inequality*

$$\langle C(F^*), F - F^* \rangle \geq \langle \Pi(q^*), q - q_* \rangle$$

holds for any feasible (F, q).

Proof:
In view of Theorem 1 (i), the proof is evident. □

The following result demonstrates the relations between generalized normal and generalized user equilibria.

Theorem 8 *Let the travel cost operator $T(x)$ be upper semicontinuous, i.e. for any feasible $\bar{x}$*

$$\liminf_{x \to \bar{x}} T(x) \geq T(\bar{x}).$$

Then any generalized normal equilibrium in the network $\mathcal{G}$ is a generalized user equilibrium.

Proof:
Let (F^*, q_*) be a generalized normal equilibrium. Then, using the same reasoning as in Theorem 4(i), we can prove that the flow pattern F^* is a user equilibrium in the network $\mathcal{G}$ with the (fixed) demand vector q_*. Let us prove now the relations (20), (21).

Let $q_*^p < \bar{q}^p$. Consider

$$q(\epsilon) = q_* + \epsilon 1_p, \quad F(\epsilon) = F^* + \epsilon 1_k^p, \quad \epsilon > 0.$$

In view of the definition of the generalized normal equilibrium, for small enough ϵ we have:

$$0 \leq \langle C(F(\epsilon)), F(\epsilon) - F^* \rangle - \langle \Pi(q(\epsilon)), q(\epsilon) - q_* \rangle = \epsilon(c_k^p(F(\epsilon)) - \Pi^p(q(\epsilon))).$$

Therefore

$$\lim_{\epsilon \to 0^+} c_k^p(F(\epsilon)) \geq \lim_{\epsilon \to 0^+} \Pi^p(q(\epsilon))$$

and we get (23).

Further, let $q_*^p > 0$. Consider

$$q(\epsilon) = q_* - \epsilon 1_p, \quad F(\epsilon) = F^* - \epsilon 1_k^p, \quad \epsilon > 0.$$

In view of the definition of the generalized normal equilibrium, for small enough ϵ we have:

$$0 \leq \langle C(F(\epsilon)), F(\epsilon) - F^* \rangle - \langle \Pi(q(\epsilon)), q(\epsilon) - q_* \rangle = \epsilon(\Pi^p(q(\epsilon)) - c_k^p(F(\epsilon))).$$

Therefore

$$\lim_{\epsilon \to 0^+} \Pi^p(q(\epsilon)) \geq \lim_{\epsilon \to 0^+} c_k^p(F(\epsilon))$$

and we get (24). □

5 Generalized Beckmann formulations

5.1 Potential operators

In this section we will show that in some cases, when the travel cost operator has additional properties concerning its "symmetry", the problem of finding a normal equilibrium can be reduced to a convex minimization problem. Since a numerical solution of a minimization problem is simpler than that of a variational inequality, it seems reasonable to study this possibility.

We start with a definition. Let $S(x) : \mathcal{D} \to E$ be an operator defined on an open convex set $\mathcal{D}$. Let x_0, $x \in Q$. Consider the function

$$f(x_0, x) = \int_0^1 \langle S(x_0 + t(x - x_0)), x - x_0 \rangle dt. \tag{26}$$

In what follows we assume that this integral does exist.

Definition 7 *We call $S(x)$ a* potential operator *if there exists a function $\phi(x)$ such that for any x_0, $x \in Q$*

$$f(x_0, x) = \phi(x) - \phi(x_0).$$

Let us present some examples of potential operators.

Lemma 1 *(i) Let $S(x)$ be differentiable and*

$$S'(x) = (S'(x))^T. \tag{27}$$

Then $S(x)$ is a potential operator.

(ii) Let $\phi(x)$ be a convex (non-smooth) function on Q. Let $S(x)$ be an element of $\partial\phi(x)$. Then $S(x)$ is a potential operator.

Proof:
(i) In view of well-known result (see, e.g. [8], page 96), in this case there exists a function $\phi(x)$ such that $f(x, y) = \phi(x) - \phi(y)$ for any x, $y \in Q$.

(ii) This part immediately follows from the identity:

$$\phi(x) - \phi(x_0) = \int_0^1 \langle S(x_0 + t(x - x_0)), x - x_0 \rangle dt.$$

□

Let us prove now a converse statement.

Theorem 9 *(i) Let $S(x)$ be a continuous potential operator. Then the function $f(x_0, x)$ is differentiable in x and $f'_x(x_0, x) = S(x)$.*

(ii) Let $S(x)$ be a potential monotone operator. Then the function $f(x_0, x)$ is convex in x and $S(x) \in \partial_x f(x_0, x)$.

Proof:
Note that for any x_0, x, $y \in \mathcal{D}$ we have:

$$f(x_0, x) + f(x, y) = \phi(x) - \phi(x_0) + \phi(y) - \phi(x) = f(x_0, y).$$

Therefore

$$f(x_0, y) - f(x_0, x) = \int_0^1 \langle S(x + t(y - x)), y - x\rangle dt. \tag{28}$$

Let $y = x + \alpha p$. Then $\frac{1}{\alpha}[f(x_0, y) - f(x_0, x)] = \int_0^1 \langle S(x + t(y - x)), p\rangle dt$. Since $S(x)$ is continuous, the limit of this expression in $\alpha \to 0$ exists and we get part (i) of the theorem.

Let us prove now part (ii) of the theorem. Since $S(x)$ is monotone, in view of (28) we have $f(x_0, y) - f(x_0, x) \geq \langle S(x), y - x\rangle$. □

Remark 1 *Note that usually the notion of the potential operator is introduced for differentiable operators by condition (27). This way is not acceptable for our goals since we work with non-differentiable and even discontinuous operators. Definition 7, which extends the standard one, does not require differentiability for $S(x)$. It is also important that potentiality of $S(x)$ sometimes can be verified analytically, since the standard travel cost functions in transportation models are very simple.*

Theorem 10 *Let $S(x)$ be a potential complete monotone operator, $x_0 \in \mathcal{D}$, and Q be a closed convex set. Assume that the set $\mathcal{F} = Q \bigcap \mathcal{D}$ is bounded. Then the set of solutions of the variational inequality*

$$\langle S(x), x - x^*\rangle \geq 0, \quad \forall x \in \mathcal{F}, \tag{29}$$

is nonempty and it coincides with the set of optimal solutions of the following convex minimization problem:

$$\min_{x \in Q} f(x_0, x), \tag{30}$$

where $f(x_0, x)$ is defined by (26).

Proof:
In view of Theorem 12, the set X^* of solutions to (29) is nonempty. For any $x^* \in X^*$ and $x \in \mathcal{F}$ we have:

$$f(x_0, x) - f(x_0, x^*) = \int_0^1 \langle S(x^* + t(x - x^*)), x - x^*\rangle dt \geq 0.$$

Thus, x^* is a solution to (30).

On the other hand, in view of Theorem 9, $f(x_0, x)$ is convex in x and $S(x) \subseteq \partial_x f(x_0, x)$. Let x^* be a solution to (30). Then there exists a vector $g^* \in \partial_x f(x_0, x^*)$ such that

$$\langle g^*, x - x^* \rangle \geq 0, \quad \forall x \in \mathcal{F}.$$

Since $\partial_x f(x_0, x)$ is a monotone operator, we have:

$$\langle S(x) - g^*, x - x^* \rangle \geq 0.$$

Thus, x^* is a solution to (29). □

5.2 Normal equilibria and minimization problems

In the previous sections we have defined different normal equilibria as solutions to corresponding variational inequalities. In the case of potential monotone travel cost operator, the solution of the variational inequality can be found from some minimization problem. The advantage of this reformulation is that the numerical methods of convex minimization are much more powerful than the numerical methods for variational inequalities.

Let us present the optimization formulation for different types of the normal equilibria. To this end let us make the following assumption.

Assumption 3 *The travel cost operator $T(x)$ is complete, monotone and potential.*

In the rest of this section we suppose that Assumption 3 holds.

Normal equilibrium with constant demand. Let $f(x) = \int_0^1 \langle T(tx), x \rangle dt$. Then in view of Theorem 10, the set of normal equilibria with constant demand coincides with the set of solution of the following minimization problem:

$$\begin{array}{ll} \min\limits_{F} & f(\Psi F) \\ \text{s.t.} & BF = q, \\ & F \geq 0. \end{array}$$

Normal equilibrium with elastic demand. Suppose the reservation cost operator $\Pi(q)$ be anti-monotone and potential. Denote

$$f(x) = \int_0^1 \langle T(tx), x \rangle dt, \quad \pi(q) = \int_0^1 \langle \Pi(tq), q \rangle dt.$$

Then in view of Theorem 10, the set of normal equilibria with elastic demand coincides with the set of solution of the following minimization problem:

$$\begin{array}{ll} \min\limits_{F} & f(\Psi F) - \pi(q) \\ \text{s.t.} & BF = q, \\ & 0 \leq q \leq \bar{q}, \\ & F \geq 0. \end{array}$$

Note that under our assumptions, $\pi(q)$ is a concave function (see Theorem 9 (ii)). Thus, the objective function in the above minimization problem is convex.

Remark 2 *The idea of finding an equilibrium flow pattern from a minimization problem was first introduced by Beckmann. He considered the simplest form of the travel cost operator $T(x) = (T_1(x_1) \dots T_A(x_A))$, which is clearly potential. Using our approach we can easily explain the objective function of the Beckmann's model:*

$$f(0, x) = \sum_{a=1}^{A} \int_0^1 T_a(tx_a) x_a dt = \sum_{a=1}^{A} \int_0^{x_a} T_a(t) dt.$$

Beckmann considered only a continuous travel cost functions. It is interesting to note that his objective function works also for discontinuous but monotone travel cost. However, in this case the solution of the corresponding optimization problem is a normal equilibrium (see theorem 10).

6 System optimum

It the previous sections we have considered several definitions of equilibrium in the network $\mathcal{G}$, which are, in fact, the variants of the user equilibrium. Let us discuss now the possibilities to compute the social optimum in the network. Note that, given a feasible flow pattern F, we can write out the social cost as follows:

$$s(F) = \langle C(F), F \rangle = \langle \Psi^T T(\Psi F), F \rangle = \langle T(x), x \rangle \equiv \hat{s}(x),$$

where $x = x(F) = \Psi F$. Thus, the problem of finding the socially optimal flow pattern F^* consists in finding the solution of the following minimization problem[2]:

$$\begin{array}{rl} \min\limits_{x,F} & \hat{s}(x) \\ \text{s.t.} & x = \Psi^T F, \\ & BF = q, \\ & F \geq 0. \end{array} \tag{31}$$

If we assume that $0 \in \mathcal{D}$, then $\hat{s}(x)$ goes to infinity as x approaches the boundary of the set $\mathcal{D}$. Therefore under this assumption the solution of (31) exists.

The main shortcoming of the problem (31) is that it is not convex even if we assume that the travel cost operator is monotone. Let us present a simple example.

Example 6 *Let the network $\mathcal{G}$ consists of one O/D pair connected by two links. Let $q = 1$ and*

$$T_1(x) = T_2(x) = 1 - e^{-ax},$$

[2] In this section we consider for simplicity only a model with inelastic demands.

where $a > 4$ is a constant. Since $T_i(x)$ is a strictly increasing function, the conditions of Theorem 2 for the existence of the unique normal equilibrium are satisfied. Moreover, in view of Theorem 4, it is also the user and the Wardrop equilibrium. Since our model is symmetric, we conclude that the equilibrium flow is given by

$$x_1^* = x_2^* = \tfrac{1}{2}.$$

Let us compute the derivatives of the social cost $\hat{s}(x)$ at this point. Since $x_1 + x_2 = 1$, we can consider $\hat{s}(x_1, x_2)$ as a function of one variable:

$$\bar{s}(y) = \hat{s}(y, 1-y) = yT_1(y) + (1-y)T_1(1-y).$$

It is easy to see that

$$\bar{s}'(y)\mid_{y=\frac{1}{2}} = 0, \quad \bar{s}''(y)\mid_{y=\frac{1}{2}} = (4a - a^2)e^{-a/2} < 0$$

since $a > 4$. Thus, we the user equilibrium in this network coincides with a local maximum of the social cost. Moreover, it is easy to see, that the social optimum is not unique: there are two points with the same minimal value of the social cost (these points are symmetric with respect to $\frac{1}{2}$).

Note that the reason for the non-uniqueness result in Example 6 is the non-convexity of the social cost. Let us write out several conditions, which are sufficient for this property.

1. Diagonal travel cost operator. Let the travel cost operator is separable:

$$T(x) = (T_1(x_1) \dots T_A(x_A)).$$

Then we need only to guarantee the convexity of functions $f_i(x) = xT_i(x)$. That will be true, for example, if we assume that all $T_i(x)$ are increasing and convex.

2. Homogeneous travel cost operator. Let the travel cost operator is homogeneous:

$$T(tx) = t^\omega T(x), \quad t \geq 0,$$

where ω is a positive constant. Then

$$f(x) = \int_0^1 \langle T(tx), x\rangle dt = \int_0^1 t^\omega dt \langle T(x), x\rangle = \tfrac{1}{1+\omega}\hat{s}(x).$$

Therefore, if $T(x)$ is a potential monotone operator, then the social cost $\hat{s}(x)$ is convex. Note that in the latter case the normal equilibrium and the social optimum coincides.

3. Linear travel cost operator. Let $T(x) = Ax$, where A is a positive-semidefinite matrix. Then the social cost is convex.

4. Constant travel cost operator. Clearly, if $T(x) \equiv d$ for some $d \in R^A$, the social cost is a sum of linear functions and therefore it is convex.

Of course, the above examples are rather restrictive. However, let us make the following observation: if we have two operators $T_1(x)$ and $T_2(x)$ such that the functions $\langle T_i(x), x\rangle$ are convex, $i = 1, 2$, then the linear combination of these operators

$$T(x) = \alpha T_1(x) + \beta T_2(x), \quad \alpha, \beta \geq 0,$$

also has this property. Thus, we can guarantee that the problem of finding a social optimum for an operator $T(x)$, formed as a linear combination of the operators mentioned in Items 1-4, can be found as a solution of a *convex* minimization problem (31).

7 Applications

7.1 Transportation Networks with Signalized Intersections

Let us start from the description of the performance of an intersection in an urban network equipped by a traffic light. Consider a node $\nu \in \mathcal{N}$ and let $I_\nu = (a_1, \ldots, a_{k_\nu})$ be the set of incoming arcs. Any schedule of this traffic light at node ν can be described by the period of the traffic light T_ν and by the vector $\Delta_\nu = \{\Delta_a, a \in I_\nu\}$ of the time periods during which the arc a_i is open.[3] Note that

$$\sum_{a\in I_\nu} \Delta_a = T_\nu.$$

Denote $\lambda_\nu = \{\lambda_a, a \in I_\nu\}$, where $\lambda_a = \Delta_a/T_\nu$. Clearly, $\lambda_a \geq 0$ and $\sum\limits_{a\in I_\nu} \lambda_a = 1$.

Let n_a be the number of drivers waiting in the queue on the arc a and $\sigma_\nu = \{n_a, a \in I_\nu\}$. Then the average waiting time on this arc can be defined as

$$W_a = \frac{n_a T_\nu}{\Delta_a s_a} = \frac{n_a}{\lambda_a s_a} \equiv W(n_a, \lambda_a),$$

where s_a is the capacity of the exit of arc a. Then the social cost of the schedule λ_ν can be defined as follows:

$$SC_\nu = SC(\lambda_\nu, \sigma_\nu) = \sum_{a\in I_\nu} W(n_a, \lambda_a) n_a = \sum_{a\in I_\nu} \frac{n_a^2}{\lambda_a s_a}.$$

Note that the function $SC(\lambda, \sigma)$ is jointly convex in (λ, σ). Moreover, the optimal schedule $\lambda(\sigma)$ can be found analytically as the solution of the following minimization problem:

$$\min\{SC(\lambda, \sigma) \mid \lambda_a \geq 0, a \in I_\nu, \sum_{a\in I_\nu} \lambda_a = 1\}.$$

Namely,

$$\lambda_a^*(\sigma_\nu) = \frac{\frac{n_a}{\sqrt{s_a}}}{\sum\limits_{b\in I_\nu} \frac{n_b}{\sqrt{s_b}}}.$$

[3] The description of more complex intersections with specific traffic lights for some lanes can be done in the similar way by introducing additional arcs associated with this node.

The minimal social cost of this intersection is thus

$$SC(\lambda^*(\sigma_\nu), \sigma_\nu) = \left[\sum_{b \in I_{nu}} \frac{n_b}{\sqrt{s_b}} \right]^2 \equiv SC^*(\sigma_\nu) \equiv SC^*_\nu.$$

Let us describe now how we can incorporate the above traffic light model in a static transportation network model. For each intersection $\nu \in \mathcal{N}$ we fix a traffic light schedule λ_ν. We assume that the individual cost for using the arc a in the route is as follows:

$$T_a(x; \lambda) = tt_a(x_a) + W(x_a, \lambda_a),$$

where $tt_a(\cdot)$ is a non-decreasing function. Then, the problem of finding the normal equilibrium is equivalent to

$$\begin{aligned} \min_F \quad & f(\Psi F; \lambda), \\ \text{s.t.} \quad & BF = q, \\ & F \geq 0, \end{aligned} \tag{32}$$

where $f(x; \lambda) = \sum\limits_{a \in \mathcal{A}} f_a(x_a, \lambda_a)$ and

$$f_a(x_a; \lambda_a) = \int\limits_0^{x_a} [tt_a(y) + W(y, \lambda_a)]dy = \int\limits_0^{x_a} tt_a(y)dy + \tfrac{x_a^2}{2\lambda_a s_a}.$$

Therefore

$$f(x; \lambda) = \sum_{a \in \mathcal{A}} \int\limits_0^{x_a} tt_a(y)dy + \tfrac{1}{2} \sum_{\nu \in \mathcal{N}} SC(\lambda_\nu, \sigma_\nu),$$

where λ is the collection of all traffic light schedules λ_ν, $\nu \in \mathcal{N}$.

Note that the objective function $f(x; \lambda)$ in the problem (32) is separable in x. However, the second term in its expression is exactly the social cost of the performance of the traffic light system. Therefore, we can assume that this system uses the optimal schedule for the current arc flow pattern x. Namely, we can substitute the optimal value $\lambda^*(\sigma_\nu)$ in the function $f(x; \lambda)$. This gives us the following expression:

$$\begin{aligned} \bar{f}(x) \quad & \equiv f(x; \lambda^*(x)) = \sum_{a \in \mathcal{A}} \int\limits_0^{x_a} tt_a(y)dy + \tfrac{1}{2} \sum_{\nu \in \mathcal{N}} SC^*(\sigma_\nu) \\ & = \sum_{a \in \mathcal{A}} \int\limits_0^{x_a} tt_a(y)dy + \tfrac{1}{2} \sum_{\nu \in \mathcal{N}} \left[\sum_{b \in I_\nu} \tfrac{x_b}{\sqrt{s_b}} \right]^2. \end{aligned}$$

Theorem 11 *The solution of the problem*

$$\begin{aligned} \min_F \quad & \bar{f}(x), \\ s.t. \quad & BF = q, \\ & x = \Psi F \\ & F \geq 0, \end{aligned} \tag{33}$$

provides us with the static user equilibrium for the network with the optimal strategy for the traffic light system.

Note that the objective function in the above problem is not separable anymore.

In our model the objective function for the social optimum is as follows

$$\hat{f}(x; \lambda) = \sum_{a \in \mathcal{A}} tt_a(x_a)x_a + \sum_{\nu \in \mathcal{N}} SC(\lambda_\nu, \sigma_\nu).$$

If we optimize the social cost in λ we get

$$\hat{f}^*(x) \equiv \hat{f}(x; \lambda^*(x)) = \sum_{a \in \mathcal{A}} tt_a(x_a)x_a + \sum_{\nu \in \mathcal{N}} SC^*(\sigma_\nu).$$

This means that we can guarantee that *the socially optimal solution can be reached by the traffic light regulation* if and only if

$$tt_a(x_a)x_a = 2 \int_0^{x_a} tt_a(y)dy.$$

This is possible if and only if $tt_a(x_a) = \gamma_a x_a$. Note that such travel time function is not realistic. The standard travel time functions are as follows:

$$tt_a(x_a) = \bar{tt}_a, , \quad tt_a(x_a) = \max\{\bar{tt}_a, \gamma_a x_a^{\beta_a}\}, \quad \text{etc.}$$

7.2 Toll policy

In the previous section we have seen that the system optimum cannot be decentralized by traffic light regulation. Let us check if this goal can be achieved with flow-dependent tolls. Such toll is just an additional cost $\tau_a(x)$, which drivers have to pay when crossing the arc a. In this situation the travel cost function for the network is as follows:

$$\tilde{T}(x) = T(x) + \tau(x),$$

where $T(x)$ is the initial travel cost and the vector $\tau(x)$ is a vector of all toll functions $\tau_a(x)$. Then the corresponding variational inequality for the normal equilibrium is as follows: find $F^* \geq 0$, $BF^* = q$ such that

$$\langle \tilde{T}(x^*), x - x^* \rangle \geq 0, \quad \forall x = \Psi F, \ BF = q, \ F \geq 0,$$

where $x^* = \Psi F^*$. On the other hand, the objective function for the social optimum is

$$f^o(x) = \langle T(x), x \rangle.$$

Let us assume that $T(x)$ is differentiable at all points of the set $S \subseteq R_+^A$ and that S is dense in R_+^A. Define the operator $T^o(x) = T'(x)x + T(x)$ all $x \in S$ For $x \notin S$ define $T^o(x) = g(x) + T(x)$, where $g(x)$ is an arbitrary point from the set

$$\mathrm{cl\,Conv}\,\{g \mid \ g = \lim_{u \to x, u \in S} T'(u)u\}.$$

Thus, the socially optimal system of tolls is defined by the equation $T^o(x) = \tilde{T}(x)$. That is

$$\tau^o(x) = \begin{cases} T'(x)x, & x \in S, \\ \\ g(x), & x \notin S. \end{cases}$$

Thus, the optimal toll is equal to the congestion externality. The user equilibrium with the optimal toll coincides with the social optimum.

For the standard diagonal travel time operator $T_a(x) = \max\{\bar{t}t_a, \gamma_a x_a^{\beta_a}\}$ the toll policy is discontinuous:

$$\tau_a^o(x) = \begin{cases} \gamma_a \beta_a x_a^{\beta_a}, & \text{if } \gamma_a x_a^{\beta_a} > \bar{t}t_a, \\ \\ 0, & \text{if } \gamma_a x_a^{\beta_a} < \bar{t}t_a. \end{cases}$$

Therefore, if we define $\tau_a^o\left(\left[\frac{\bar{t}t_a}{\gamma_a}\right]^{1/\beta_a}\right) = \bar{t}t_a$, we can guarantee that any normal equilibrium will be a user equlibrium (see Theorem 4(i)).[4] Note that the travel cost operator with toll is therefore also *discontinuous*.

7.3 Limited capacities

Consider a model of the travel time/density relationships described by the following parameters:

- L - length of the arc,
- l - average length of a car,
- δ - average reaction time of a driver.

We use the following assumption:

> *The distance Δ between two consecutive cars is sufficient to avoid an accident if the leading car suddenly starts to break upto a full stop.*

[4]See also Section 6 for discussion the cases when the problem of finding the social optimum is a variational inequality with a monotone operator.

Let a be an average performance of the breaks. Denote by $x_i(t)$, $i = 1, 2$, the positions of the front wheels of two consecutive cars on the road. Assume that at time $t = 0$ both cars keep the same speed v. At this moment the first car starts to break. Denote by S the total distance covered by the car before it stops: $x_1(v/a) = x_1(0) + S$. The second car reacts with a delay δ (*reaction time*). Therefore it stops at the position

$$x_2(\delta + v/s) = x_2(0) + \delta v + S.$$

Since during the period $0 \leq t \leq \delta + v/a$ we always have $x_1'(t) \leq x_2'(t)$, the safety distance is

$$\Delta \geq \delta v + l.$$

If x drivers are distributed uniformly on the road, then $\Delta \leq \frac{L}{x}$. Therefore

$$v \leq \frac{1}{\delta}\left(\frac{L}{x} - l\right).$$

Hence, we have the following lower estimate for the travel time on this road:

$$tt(x) \geq \frac{\delta L x}{L - lx}.$$

Since our reasoning assumes congestion, the final form of the travel time function is

$$tt(x) = \max\left\{\bar{tt}, \frac{\delta L x}{L - lx}\right\},$$

where $\bar{tt}$ is a free traffic travel time. Note that in accordance to this expression the switching load is defined as

$$\bar{x} = \frac{\bar{tt} L}{\delta L + \bar{tt} l} = \frac{L}{l + \delta \bar{v}},$$

where $\bar{v}$ is the speed limit.

The proposed travel time function justifies our interest to unbounded monotone operators. Note that it is increasing and convex. The Beckmann form of function $tt(x)$, which is used for finding an equilibrium by a minimization problem, is as follows:

$$f(x) = \int\limits_0^x \max\left\{\bar{tt}, \tfrac{\delta L \tau}{L - l\tau}\right\} d\tau.$$

It is interesting that the form of $tt(x)$ is suitable for finding a system optimum since the function $x tt(x)$ is convex.

8 Conclusion

In this paper we have considered different formulations for finding an equilibrium in static congested transportation networks. All these models were written in terms of path flow. Note that the dimension of this vector even for a small network comprising several dozen of nodes can be beyond of the abilities of the most powerful computer. Therefore, these formulations in their current form are computationally intractable (except if complete path enumeration is avoided using a heuristic rule based on some notion of reasonable paths). However, it can be shown that they can be rewritten as some problems in much smaller dimension. This and other computational aspects will be discussed in a forthcoming paper.

References

[1] Arnot, R., Arrow, K., Atkinson, A. and Drèze, J. *Public Economics: Selected Papers by W. Vickrey.* Cambridge University Press, Cambridge, 1994.

[2] Beckmann, M., C. McGuire and C. Winsten. *Studies in Economics of Transportation*, Yale University Press, New Haven, 1956.

[3] Bernstein, D. and T.E. Smith. "Equilibria for Networks with Lower Semicontinuous Costs: With an Application to Congested Pricing". *Transportation Science*, 28, pp. 221–235, 1994.

[4] Dafermos, S., "Traffic Equilibrium and Variational Inequalities." *Transportation Science*, 14, pp. 42 – 54, 1980.

[5] Kinderlehrer, D., and Stampacchia, G. *An Introduction to Variational Inequalities and Their Applications.* Academic Press, New York, 1980.

[6] Nagurney., A. *Network Economics: A Variational Inequality Approach.* Kluwer Academic Publishers, Dordrecht, 1993.

[7] Nesterov, Yu., and Nemirovsky, A. *Interior Point Polynomial Algorithms in Nolinear Optimization.* SIAM, Philadelphia, 1994.

[8] Ortega, J.M., and Rheinboldt, W.C. *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, New York, 1970.

[9] Sheffi, Y. *Urban Transportation Networks: Equlibrium Analysis with Mathematical Programming Methods.* Prentice-Hall, Englewood Cliffs, New Jersey, 1985.

[10] Smith, M.J. "Existence, Uniqueness, and Stability of Traffic Equilibria". *Transportation Research*, 13B, pp. 259 – 304, 1979.

[11] Wardrop, J.G. "Some Theoretical Aspects of Road Traffic Research", in *Proceedings of the Institute of Civil Engineers,* Part II, pp. 325-378, 1952.

9 Appendix. Variational inequalities

In this section we give a brief description of the notion of variational inequality and present some results which are used in our analysis of equilibrium. For completeness of presentation we provide almost all of them with a simple proof, especially those related to the notion of *complete* monotone operator. For standard results and references see, e.g. [6], Chapter 1.

Let E be a finite dimensional real vector space and let S be a multi-valued mapping defined on an open convex set $\mathcal{D} \subseteq E$ and taking values in E. More precisely, using S we associate with any point $x \in \mathcal{D}$ a nonempty subset $S(x) \subset E$. If $S(x)$ is single-valued (i.e. $S(x)$ consists of a single point for every $x \in \mathcal{D}$), then the (unique) point of the set $S(x)$, $x \in \mathcal{D}$, is also denoted by $S(x)$.

Definition 8 *The mapping S is called* monotone *if for any $x_1 \in \mathcal{D}$, $g_1 \in S(x_1)$, $x_2 \in \mathcal{D}$ and $g_2 \in S(x_2)$*

$$\langle g_1 - g_2, x_1 - x_2 \rangle \geq 0.$$

If for $x_1 \neq x_2$ the above inequality is always strict, the mapping $S(x)$ is called strictly monotone. *If $-S(x)$ is monotone, we call $S(x)$ the* antimonotone *mapping.*

Note that the monotonicity is invariant with respect to affine substitution of variables: if $S(x)$ is a monotone mapping defined on a convex set $\mathcal{D}$, and $A(y)$ is a linear operator then the mapping $\hat{S}(y) = A^T S(A(y))$ is monotone on the set $\hat{\mathcal{D}} = \{y \mid A(y) \in \mathcal{D}\}$. Indeed, if $y_1, y_2 \in \hat{\mathcal{D}}$ then there exist $x_i = A(y_i) \in \mathcal{D}$ and therefore

$$\langle \hat{S}(y_1) - \hat{S}(y_2), y_1 - y_2 \rangle = \langle S(x_1) - S(x_2), x_1 - x_2 \rangle \geq 0.$$

Monotonicity is also preserved by summation: if mappings $S_1(x)$ and $S_2(x)$ are monotone then the mapping $\alpha S_1(x) + \beta S_2(x)$ with $\alpha, \beta \geq 0$, is also monotone.

Example 7 *Let $S(x)$ be a differentiable mapping such that for any $x \in \mathcal{D}$ the matrix $S'(x)$ is positive semidefinite. Then the mapping $S(x)$ is monotone. Moreover, if $S'(x)$ is positive definite, then $S(x)$ is strictly monotone.* The proof is immediate since

$$\langle S(x) - S(y), x - y \rangle = \int_0^1 \langle S'(y + \theta(x-y))(x-y), x - y \rangle d\theta. \quad \square$$

Example 8 *Let $S(x)$ be a diagonal dominant differentiable mapping: for any $x \in \mathcal{D}$*

$$S'(x)_{ii} \geq \tfrac{1}{2} \sum_{j \neq i} \mid S'(x)_{ij} + S'(x)_{ji} \mid, \quad i = 1 \ldots n.$$

Then the mapping $S(x)$ is monotone.
Proof:

In view of the previous example, we shall prove only that the matrix $S'(x)$ is positive semidefinite. Note that without loss of generality we can assume that $S'(x)$ is symmetric. Let λ be an eigenvalue of the matrix and y be the corresponding nonzero eigenvector: $S'(x)y = \lambda y$. Let k be the index of the component of y with the largest absolute value. Then

$$| S'(x)_{kk} - \lambda | \cdot | y^{(k)} | = | \sum_{j \neq k} S'(x)_{kj} y^{(j)} | \leq | (S'(x))_{kk} | \cdot | y^{(k)} | .$$

Hence, any eigenvalue of $S'(x)$ is non-negative, or, $S'(x)$ is positive semidefinite. □

Example 9 *Let $V(x)$ be a contraction mapping:*

$$\| V(x) - V(y) \| \leq q \| x - y \|, \quad \forall x, y \in \mathcal{D},$$

where the constant $q \in [0, 1]$. Then the operator $S(x) = x - V(x)$ is monotone:

$$\langle S(x) - S(y), x - y \rangle = \| x - y \|^2 - \langle V(x) - V(y), x - y \rangle \geq (1 - q) \| x - y \|^2 .$$

If $q < 1$, then $S(x)$ is strictly monotone.

In our analysis of the equilibrium we need the following simple fact.

Lemma 2 *Let a single-valued mapping $S(x)$ be monotone, $x \in \mathcal{D}$ and $v \in E$. Then for points $x(\epsilon) = x + \epsilon v$ the limit*

$$\lim_{\epsilon \to 0^+} \langle S(x(\epsilon)), v \rangle \tag{34}$$

does exist.

Proof:
Let $\xi(\epsilon) = \langle S(x(\epsilon)), v \rangle$, $\epsilon > 0$. Then for any $\epsilon_1 > \epsilon_2 > 0$ such that $x(\epsilon_1), x(\epsilon_2) \in \mathcal{D}$ we have:

$$0 \leq \langle S(x(\epsilon_1)) - S(x(\epsilon_2)), x(\epsilon_1) - x(\epsilon_2) \rangle = (\epsilon_1 - \epsilon_2) \langle S(x(\epsilon_1)) - S(x(\epsilon_2)), v \rangle.$$

This implies that the function $\xi(\epsilon)$ is non-decreasing in ϵ and below bounded (by $\xi(0)$). Therefore the limit (34) exists. □

Let Q be a convex set. Denote $\mathcal{F} = Q \bigcap \mathcal{D}$.

Definition 9 *The* variational inequality *problem consists in finding a point $x^* \in \mathcal{F}$ such that for any $x \in \mathcal{F}$ and $g \in S(x)$ we have*

$$\langle g, x - x^* \rangle \geq 0. \tag{35}$$

The point x^* is called a solution to the variational inequality (35). If there exists some $g^* \in S(x^*)$ such that

$$\langle g^*, x - x^* \rangle \geq 0 \tag{36}$$

for all $x \in \mathcal{F}$, the pair (x^*, g^*) is called the strong *solution to (35).*

In view of monotonicity of $S(x)$ the existence of a strong solution (x^*, g^*) implies that x^* is a solution to (35). The converse is not true: The optimal g^* can be absent in $S(x^*)$.

Lemma 3 *If $S(x)$ is single-valued and continuous then any solution x^* to (35) is strong with $g^* = S(x^*)$.*

Proof:
Let $x_0 \in \mathcal{F}$. Consider $x_\alpha = x_0 + \alpha(x^* - x_0)$. Then for any $\alpha \in [0,1]$ we have:

$$0 \leq \langle S(x_\alpha), x_\alpha - x^* \rangle = (1-\alpha)\langle S(x_\alpha), x_0 - x^* \rangle.$$

Thus, $\langle S(x_\alpha), x_0 - x^* \rangle \geq 0$. Taking the limit as $\alpha \to 1$ we get $\langle S(x^*), x_0 - x^* \rangle \geq 0$ for any $x_0 \in \mathcal{F}$. $\square$

Example 10 *Let $f(x)$, $x \in E$, be a convex (nonsmooth) function. Then the* subdifferential *$\partial f(x)$ of f at x is defined as follows:*

$$\partial f(x) = \{g \in E \mid f(y) \geq f(x) + \langle g, y - x \rangle, \ \forall y \in E\}.$$

It is well known that the mapping $S(x) = \partial f(x)$ is monotone. Note that if $f(x)$ is differentiable then $S(x)$ is single-valued: $\partial f(x) \equiv \{f'(x)\}$.

For $S(x) = \partial f(x)$ the variational inequality problem (35) consists in finding a point x^* such that

$$\langle g, x - x^* \rangle \geq 0 \quad \forall x \in Q, \ g \in \partial f(x).$$

This means that we are looking for a constrained minimum of function $f(x)$ over the set Q. If $Q \equiv E$, we will find a point x^* such that $0 \in \partial f(x^*)$. In the latter case, if $f(x)$ is differentiable, we will have $f'(x^*) = 0$.

Example 11 *Let a single-valued mapping $V(x)$, $x \in E$, be a contraction. We are looking for its fixed point: $x^* = V(x^*)$. Then this problem is clearly equivalent to (35) with $S(x) = x - V(x)$. Since we have no constraints now, (35) is satisfied if and only if $S(x^*) = 0$.*

Let us describe the structure of the solutions to (35).

Lemma 4 *The set X^* of the solutions to variational inequality (35) is convex. If $S(x)$ is strictly monotone and $X^* \neq \emptyset$ then the solution is unique.*

Proof:
Indeed, X^* is convex as an intersection of half-spaces. Let $S(x)$ be strictly monotone and $X^* \neq \emptyset$. Assume that two points x_0, $x_1 \in X^*$ are different. Then for $x_\alpha = \alpha x_1 + (1-\alpha)x_0 \in X^*$ and $g_\alpha \in S(x_\alpha)$, $\alpha \in (0,1)$, we have:

$$\begin{aligned} 0 &\leq \langle g_\alpha, x_\alpha - x_0 \rangle = \alpha \langle g_\alpha, x_1 - x_0 \rangle, \\ 0 &\leq \langle g_\alpha, x_\alpha - x_1 \rangle = (1-\alpha)\langle g_\alpha, x_0 - x_1 \rangle. \end{aligned}$$

Therefore $\langle g_\alpha, x_1 - x_0\rangle = 0$. On the other hand, for arbitrary α, $\beta \in (0,1)$ we have:

$$0 < \langle g_\alpha - g_\beta, x_\alpha - x_\beta\rangle = (\alpha - \beta)\langle g_\alpha, x_1 - x_0\rangle + (\beta - \alpha)\langle g_\beta, x_1 - x_0\rangle = 0.$$

This contradiction proves the lemma. $\square$

Let us study now the existence of the solution to the variational inequality (35). We start from a general result.

Lemma 5 *If $S(x)$ is a monotone operator and $\mathcal{F}$ is bounded then there exists a point $x_* \in \mathrm{cl}\,(\mathcal{F})$ satisfying the condition (35).*

Proof:
Consider the following system of closed convex sets:

$$Q(x,g) = \{y \in \mathrm{cl}\,(\mathcal{F}) :\ \langle g, x - y\rangle \geq 0\ \}, \quad x \in \mathcal{F},\ g \in S(x).$$

Assume first that for any set of $n+1$ pairs $(x_1, g_1), \ldots, (x_{n+1}, g_{n+1})$ with $x_i \in \mathcal{F}$ and $g_i \in S(x_i)$ the intersection $\bigcap\limits_{i=1}^{n+1} Q(x_i, g_i)$ is not empty. Then, in view of Helly theorem there exists a point

$$y_* \in \bigcap_{x \in \mathcal{F},\, g \in S(x)} Q(x,g).$$

This point clearly satisfies conditions (35).

Assume now that there exists a set of pairs $(x_1, g_1), \ldots, (x_{n+1}, g_{n+1})$, $x_i \in \mathcal{F}$, $g_i \in S(x_i)$, for which $\bigcap\limits_{i=1}^{n+1} Q(x_i, g_i) = \emptyset$. Consider the following function:

$$\phi(y) = \max_{1 \leq i \leq n+1} \langle g_i, y - x_i\rangle.$$

In view of our last assumption

$$\phi_* \equiv \min_{y \in \mathrm{cl}\,(\mathcal{F})} \phi(y)\ > 0.$$

Denote $y_* = \arg \min\limits_{y \in \mathrm{cl}\,(\mathcal{F})} \phi(y)$. Then there exists some $g_* \in \partial\phi(y_*)$ such that

$$\langle g_*, y - y_*\rangle \geq 0 \quad \forall y \in \mathrm{cl}\,(\mathcal{F}).$$

Note that

$$\partial\phi(y_*) = \mathrm{Conv}\,\{g_i,\ i \in I_*\}, \quad I_* = \{i :\ \langle g_i, y_* - x_i\rangle = \phi_*\}.$$

Thus, $g_* = \sum\limits_{i \in I_*} \alpha_i g_i$ for some $\alpha_i > 0$, $\sum\limits_{i \in I_*} \alpha_i = 1$. Therefore

$$\phi_* = \sum_{i \in I_*} \alpha_i \langle g_i, y_* - x_i\rangle = \langle g_*, y_*\rangle - \sum_{i \in I_*} \alpha_i \langle g_i, x_i\rangle.$$

On the other hand, $\langle g_i - g_j, x_i - x_j\rangle \geq 0$. Multiplying this inequality by $\alpha_i\alpha_j$ and taking the sum in all i, j we get the following:

$$\begin{aligned} 0 &\leq \sum_{i\in I_*}\sum_{j\in I_*} \alpha_i\alpha_j\langle g_i - g_j, x_i - x_j\rangle \\ &= 2\sum_{i\in I_*} \alpha_i\langle g_i, x_i\rangle - 2\langle \sum_{i\in I_*} \alpha_i g_i, \sum_{i\in I_*} \alpha_i x_i\rangle \\ &= 2\sum_{i\in I_*} \alpha_i\langle g_i, x_i\rangle - 2\langle g_*, \hat{y}\rangle, \end{aligned}$$

where $\hat{y} = \sum_{i\in I_*} \alpha_i x_i$. Thus, $0 < \phi_* \leq \langle g_*, y_* - \hat{y}\rangle \leq 0$ since $\hat{y} \in \mathcal{F}$. We get a contradiction which proves the lemma. □

In simple situation when a closed bounded set Q belongs to the domain $\mathcal{D}$, Lemma 5 implies the existence of the solution of the variational inequality problem (35). However, in many interesting cases the set $\mathcal{F} = Q \bigcap \mathcal{D}$ is not closed. Then, in order to ensure the existence of the solution we need to introduce some assumptions on the operator $S(x)$. A natural assumption of that type can be formulated in terms of *complete* operators.

Definition 10 *An operator $S(x)$, $x \in \mathcal{D}$ is called* complete *if for any $y \in \mathcal{D}$ and any sequence $\{x_k\}_{k=0}^{\infty} \subset \mathcal{D}$ converging to a boundary point of the set $\mathcal{D}$ we have*

$$\lim_{k\to\infty} \langle g_k, x_k - y\rangle = +\infty,$$

where g_k are arbitrary vectors from $S(x_k)$.

Such operators can be used for description of the equilibrium models, in which the cost goes to infinity as the argument approaches the boundary of the domain. Note that the sum of complete operators with positive coefficients is a complete operator which domain is the intersection of the initial domains.

Example 12 *Let $f(x)$ be a convex* barrier *function for an open convex set $\mathcal{D}$. This means that for any sequence $\{x_k\}_{k=0}^{\infty} \subset \mathcal{D}$ converging to a boundary point of $\mathcal{D}$ the sequence of function values goes to $+\infty$. Then $S(x) = \partial f(x)$ is a complete operator: for $y \in \mathcal{D}$ and any $g_k \in \partial f(x_k)$ we have:*

$$\langle g_k, x_k - y\rangle \geq f(x_k) - f(y) \;\to\; +\infty$$

as the sequence $\{x_k\}$ converges to the boundary of $\mathcal{D}$.

Lemma 6 *Let $S(x)$, $x \in \mathcal{D}$, be a complete monotone operator. If (x^*, g^*) is a strong solution to the variational inequality (35) then $\langle g^*, x - x^*\rangle \geq 0$ for any $x \in Q$.*

Proof:

Let us assume that there exists $y \in Q$ such that $\langle g^*, y - x^* \rangle < 0$. Since $\mathcal{D}$ is an open set and $x^* \in \mathcal{D}$, for small enough $\alpha \in (0,1)$ we have

$$y_\alpha = \alpha y + (1-\alpha)x^* \in \mathcal{D}.$$

Thus, $y_\alpha \in Q \bigcap \mathcal{D}$. However, $\langle g^*, y_\alpha - x^* \rangle = \alpha \langle g^*, y - x^* \rangle < 0$ and (x^*, g^*) cannot be a strong solution to (35). $\square$

Thus, a strong solution x^* is a minimum of the linear function $\langle g^*, x \rangle$ over the convex set Q. Note that such description does not include explicitly the domain $\mathcal{D}$.

Let us present an existence theorem for variational inequality problem (35).

Theorem 12 *Assume that the set Q is convex and closed, the operator $S(x)$ is monotone and complete and the intersection $Q \bigcap \mathcal{D}$ is bounded. Then:*

1. *There exists a solution to variational inequality (35).*
2. *If $S(x)$ is single-valued and continuous any solution to (35) is strong.*

Proof:
1. In view of Lemma 5 there exists a point $x^* \in \mathrm{cl}\,(\mathcal{F})$ such that

$$\langle g, x - x^* \rangle \geq 0$$

for all $x \in \mathcal{F}$ and $g \in S(x)$. Let us choose an arbitrary $x_0 \in \mathcal{F}$ and consider the points $x_\alpha = x_0 + \alpha(x^* - x_0)$, $g_\alpha \in S(x_\alpha)$, $\alpha \in (0,1)$. Then

$$0 \geq \langle g_\alpha, x^* - x_\alpha \rangle = (1-\alpha)\langle g_\alpha, x^* - x_0 \rangle = \tfrac{1}{\alpha}(1-\alpha)\langle g_\alpha, x_\alpha - x_0 \rangle.$$

Thus, $0 \geq \langle g_\alpha, x_\alpha - x_0 \rangle$ for all $\alpha \in (0,1)$. Note that $x_0 \in \mathcal{D}$ and $x_\alpha \to x^*$ as $\alpha \to 1$. Since $S(x)$ is complete, we conclude that $x^* \in \mathcal{D}$.

2. This statement follows from the previous item and Lemma 3. $\square$