Sustainable Transportation Networks: Foundations: Part I

Anna Nagurney
Isenberg School of Management
University of Massachusetts
Amherst, MA 01003

Transportation Networks

Transportation networks are complex, typically large-scale systems and the study of their efficient operation, often through some outside intervention, has attracted much interest from economists, engineers, as well as transportation and urban planners and operations researchers.

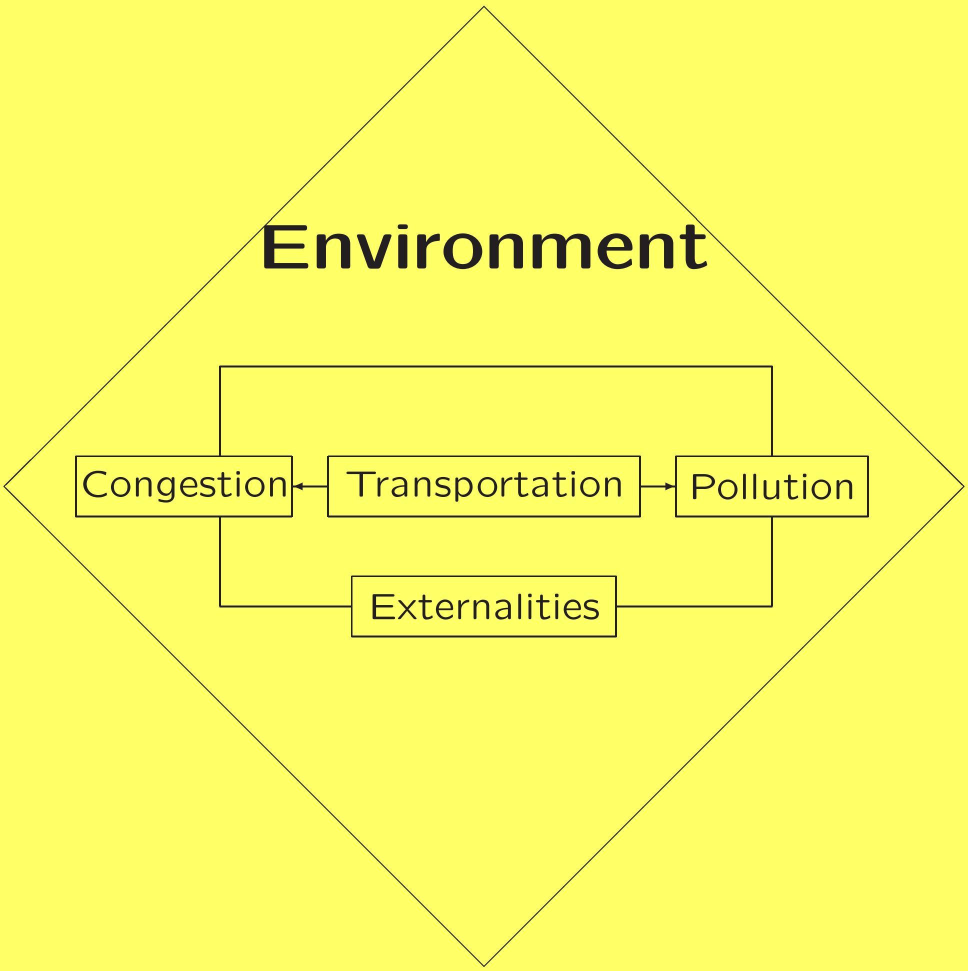
The subject dates to ancient times with such classical examples including the publicly provided Roman road network and the "time of day" chariot policy, whereby chariots were banned from the ancient city of Rome at particular times of the day (see Banister and Button 1993).

Early Contributors

From an economic perspective, some of the earliest contributions to the subject date to Pigou (1920), who is also viewed as the forefather of road pricing. In particular, his idea of using road pricing to regulate traffic congestion on a simple two-node, two-link network has spawned additional research, discussion, and practical applications.

Pigou (1920) (see also Knight 1924) used the example of a congested road to illustrate the concepts of externality and optimal tolls as congestion charges.

Specifically, Pigou argued that travelers should be charged according to their marginal external congestion costs.



Some externality relationships in a transportation system

Negative Externalities

Note that an *externality* is present when the actions of some economic agents (such as travelers) affect the utility (typically travel time or cost in the case of transportation) or production set of another without that person's consent or compensation.

An externality is seen as *negative* when the harm done to others is considered to be uncompensated. In the case of transportation networks, the harm may include, for example, increased travel time due to congestion, or increased pollution. See the Figure for a depiction of externality relationships in a transportation system.

Earlier in 1844, Dupuit had used a bridge as an illustration of the concept of efficient pricing of public goods.

Property Rights

Coase (1960), in turn, utilized a transportation example, among others, in the form of sparks from a railway, when he addressed the absence of property rights in relation to the existence of externalities. *Property rights* refer to the group of entitlements comprising the property owner's rights and privileges plus the constraints imposed for the use of the resource.

Property rights are typically distinguished as to whether they are: *private*, that is, held by individuals or firms; *common*, that is, held by an identifiable group; *state*, that is, held by the government; or of *open access*, in which no explicit ownership applies and, hence, the property is open or available to all.

“Cost of Congestion”

As early as the 1960s, Vickrey (1960, 1963) emphasized that the cost of congestion was high, with the real economic cost of transportation infrastructure in the United States at that time being estimated to be approximately three times the total vehicular and gasoline taxes generated by car use on urban streets.

Moreover, he argued in 1959 (see Hau 1998) that the results of not charging travelers for their rush-hour usage could be “disastrously expensive.”

Wardrop's Principles

Engineers were also concerned about the operation of transportation networks. In particular, Wardrop (1952) explicitly recognized alternative possible behaviors of users of transportation networks and stated two principles, which are commonly named after him:

First Principle: The journey times of all routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route.

Second Principle: The average journey time is minimal.

The first principle corresponds to the behavioral principle in which travelers seek to (unilaterally) determine their minimal costs of travel whereas the second principle corresponds to the behavioral principle in which the total cost in the network is minimal.

First Rigorous Formulations of Traffic Network Equilibrium

Beckmann, McGuire, and Winsten (1956) were the first to rigorously formulate these conditions mathematically, as had Samuelson (1952) in the framework of spatial price equilibrium problems in which there were, however, no congestion effects.

Specifically, Beckmann, McGuire, and Winsten (1956) established the equivalence between the *traffic network equilibrium* conditions, which state that all used paths connecting an origin/destination pair will have equal and minimal travel times (or costs) (corresponding to Wardrop's first principle), and the Kuhn-Tucker conditions of an appropriately constructed optimization problem, under a symmetry assumption on the underlying functions.

Hence, in this case, the equilibrium link and path flows could be obtained as the solution of a mathematical programming problem. Their approach made the formulation, analysis, and subsequent computation of solutions to traffic network problems based on actual transportation networks realizable.

User-Optimization versus System-Optimization

Dafermos and Sparrow (1969) coined the terms *user-optimized* (U-O) and *system-optimized* (S-O) transportation networks to distinguish between two distinct situations in which, respectively, users act unilaterally, in their own self-interest, in selecting their routes, and in which users select routes according to what is optimal from a societal point of view, in that the total cost in the system is minimized.

In the latter problem, marginal costs rather than average costs are equilibrated.

The former problem coincides with Wardrop's first principle, and the latter with Wardrop's second principle.

See the Table for the two distinct behavioral principles underlying transportation networks. The concept of “system-optimization” is also relevant to other types of “routing models” in transportation (as well as in communications), including those concerned with the routing of freight.

Dafermos and Sparrow (1969) also provided explicit computational procedures, that is, *algorithms*, to compute the solutions to such network problems in the case where the user travel cost on a link was an increasing (in order to handle congestion) function of the flow on the particular link and linear.

Distinct behavior on transportation networks

User-Optimization	System-Optimization
$\Downarrow$	$\Downarrow$
Equilibrium Principle:	Optimality Principle:
User travel costs on used paths for each O/D pair are equalized and minimal.	Marginals of the total travel cost on used paths for each O/D pair are equalized and minimal.

For definiteness, and for easy reference, both the classical system-optimized traffic network model and then the classical user-optimized network model are presented.

Subsequently, I discuss toll policies which guarantee that the user-optimized flow pattern is also system-optimizing. Note that at this point in the discussion in this chapter no environmental concerns have as yet been raised. I later return to this topic in in this lecture.

I also consider more general models, in which the user link travel cost functions are no longer separable and are also asymmetric. I construct link- and path-toll policies for such transportation networks. Subsequently, I then consider the U-O traffic network problem with general user link travel cost functions and provide the variational inequality formulations of the governing equilibrium conditions, since, in this case, the conditions can no longer be reformulated as the Kuhn-Tucker conditions of a convex optimization problem. Finally, I present the variational inequality formulations in the case of elastic travel demands.

This lecture, subsequently, focuses on transportation and the environment, and describes some basic policy instruments drawn from environmental economics.

I also present highlights of some unique characteristics of transportation networks as regards environmental issues.

System-Optimization Versus User-Optimization

Now, the basic traffic network models are reviewed, under distinct assumptions of their operation and distinct behavior of the travelers or users of the network. The models are classical and due to Beckmann, McGuire, and Winsten (1956) and Dafermos and Sparrow (1969).

System-Optimization

Consider a general network $G = [N, L]$, where N denotes the set of nodes, and L the set of directed links. Let a denote a link of the network connecting a pair of nodes, and let p denote a path consisting of a sequence of links connecting an O/D pair. In transportation networks, nodes correspond to origins and destinations, as well as to intersections. Links, on the other hand, correspond to roads/streets. A path, thus, is a sequence of roads which comprise a route from an origin to a destination. P_w denotes the set of paths connecting the origin/destination (O/D) pair of nodes w. Let P denote the set of all paths in the network and assume that there are J origin/destination pairs of nodes.

Let x_p represent the flow on path p and let f_a the load on link a. The path flows on the network are grouped in the column vector $x \in R_+^{n_P}$, where n_P denotes the number of paths in the network. The link loads, in turn, are grouped into the column vector $f \in R_+^n$, where n denotes the number of links in the network.

The Conservation of Flow Equations

The following conservation of flow equation must hold:

$$f_a = \sum_{p \in P} x_p \delta_{ap}, \quad \forall a \in L, \tag{1}$$

where $\delta_{ap} = 1$, if link a is contained in path p, and 0, otherwise. Expression (1) states that the load on a link a is equal to the sum of all the path flows on paths p that contain (traverse) link a.

Moreover, if one lets d_w denote the demand associated with O/D pair w, then one must have that

$$d_w = \sum_{p \in P_w} x_p, \quad \forall w \in W, \tag{2}$$

where $x_p \geq 0$, $\forall p \in P$; that is, the sum of all the path flows between an origin/destination pair w must be equal to the given demand d_w.

The User Cost Functions

Let c_a denote the user link travel cost associated with traversing link a, and let C_p denote the user cost associated with traversing the path p.

Assume that the user link travel cost function is given by

$$c_a = c_a(f_a), \quad \forall a \in L, \tag{3}$$

where c_a is assumed to be an increasing function of the link load f_a in order to model the effect of the link load on the travel cost.

The total cost on link a, denoted by $\hat{c}_a(f_a)$, hence, is given by:

$$\hat{c}_a(f_a) = c_a(f_a) \times f_a, \quad \forall a \in L, \tag{4}$$

that is, the total cost on a link is equal to the user link travel cost on the link times the flow on the link. Here the travel cost is interpreted in a general sense but often principally from a transportation engineering perspective, the travel cost on a link is assumed to coincide with the travel time on a link.

See the Figure for such a typical user link travel time function, where the free flow travel time refers to the travel time to traverse a link when there is zero load on the link (or zero vehicles).

In the system-optimized problem, there exists a central controller who seeks to minimize the total cost in the network system, where the total cost is expressed as

$$\sum_{a \in L} \hat{c}_a(f_a), \tag{5}$$

where the total cost on a link is given by expression (4).

The System-Optimized Problem

The system-optimization problem is, thus, given by:

$$\text{Minimize} \quad \sum_{a \in L} \hat{c}_a(f_a) \tag{6}$$

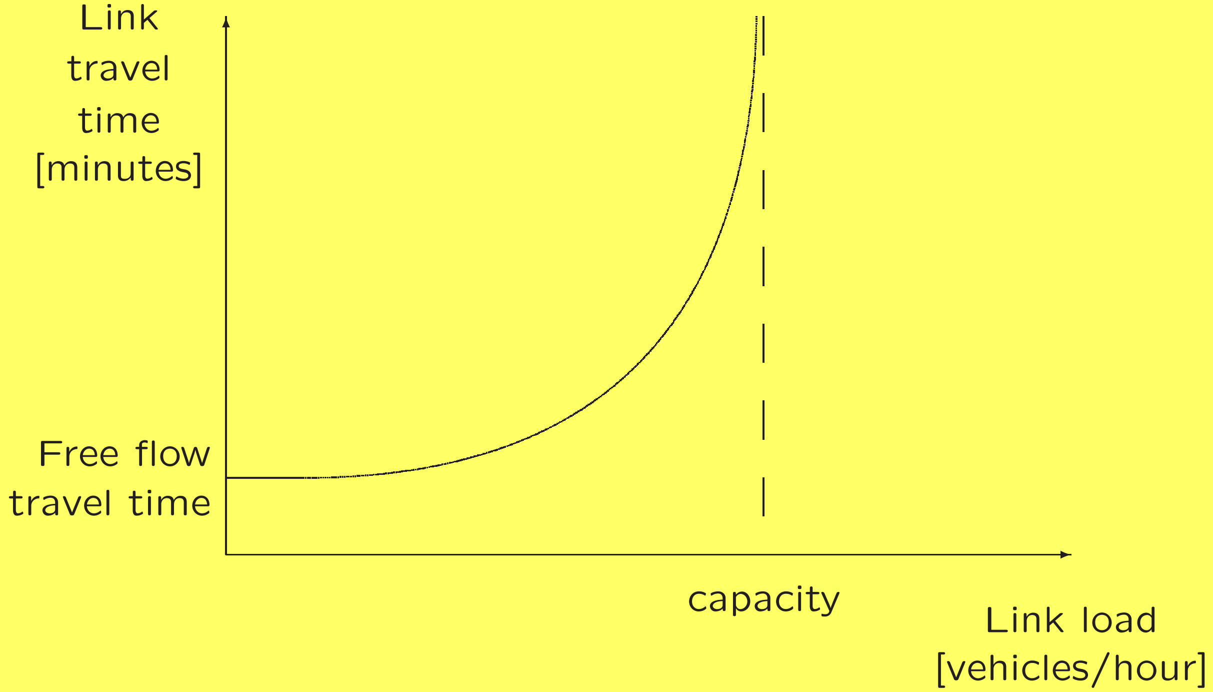
subject to:

$$\sum_{p \in P_w} x_p = d_w, \quad \forall w \in W, \tag{7}$$

$$f_a = \sum_{p \in P} x_p, \quad \forall a \in L, \tag{8}$$

$$x_p \geq 0, \quad \forall p \in P. \tag{9}$$

The constraints (7) and (8), along with (9), are commonly referred to in network terminology as *conservation of flow equations*. In particular, they guarantee that the flow in the network, that is, the travelers, do not "get lost."



A graph of a typical user link travel time function

The total cost on a path, denoted by $\hat{C}_p$, is the user link travel cost on a path times the flow on a path, that is,

$$\hat{C}_p = C_p x_p, \quad \forall p \in P, \tag{10}$$

where the user travel cost on a path, C_p, is given by the sum of the user link travel costs on the links that comprise the path, that is,

$$C_p = \sum_{a \in L} c_a(f_a)\delta_{ap}, \quad \forall a \in L. \tag{11}$$

In view of (8), one may express the cost on a path p as a function of the path flow variables and, hence, an alternative version of the above system-optimization problem can be stated in path flow variables only, where one has now the problem:

$$\text{Minimize} \quad \sum_{p \in P} C_p(x)x_p \tag{12}$$

subject to constraints (7) and (9).

Hence, without any loss of generality, I express the travel cost on a path as a function of either the path flow or the link load pattern.

System-Optimality Conditions

Under the assumption of increasing user link cost functions, the objective function in the S-O problem is convex, and the feasible set consisting of the linear constraints is also convex. Therefore, the optimality conditions, that is, the Kuhn-Tucker conditions are: For each O/D pair $w \in W$, and each path $p \in P_w$, the flow pattern x (and link load pattern f), satisfying (7)–(9) must satisfy:

$$\hat{C}'_p \begin{cases} = \mu_w, & \text{if} \quad x_p > 0 \\ \geq \mu_w, & \text{if} \quad x_p = 0, \end{cases} \tag{13}$$

where $\hat{C}'_p$ denotes the marginal of the total cost on path p, given by:

$$\hat{C}'_p = \sum_{a \in L} \frac{\partial \hat{c}_a(f_a)}{\partial f_a} \delta_{ap}, \tag{14}$$

and in (13) is evaluated at the solution.

In the S-O problem, it is the marginal of the total costs on each used path connecting an O/D pair which are equalized and minimal. Conditions (13) state that a system-optimized flow pattern is such that for each origin/destination pair the incurred marginals of the total cost on each used path are equal and minimal.

User-Optimization

Now, the user-optimized traffic network problem is considered, also commonly referred to as the *traffic assignment* problem or the *traffic network equilibrium* problem. Again, as in the system-optimized problem, the network $G = [N, L]$, the travel demands associated with the origin/destination pairs, as well as the user link travel cost functions are assumed as given. Recall that user-optimization follows Wardrop's first principle.

Traffic Network Equilibrium Conditions

Now, however, one seeks to determine the path flow pattern x^* (and link load pattern f^*) which satisfies the conservation of flow equations (1), (2) and the nonnegativity assumption on the path flows, and which also satisfies the traffic network equilibrium conditions given by the following statement.

For each O/D pair $w \in W$ and each path $p \in P_w$:

$$C_p \begin{cases} = \lambda_w, & \text{if} \quad x_p^* > 0 \\ \geq \lambda_w, & \text{if} \quad x_p^* = 0. \end{cases} \tag{15}$$

Hence, in the user-optimization problem there is no explicit optimization concept, since now travelers act independently, in a noncooperative manner, until they cannot improve on their situations unilaterally and, thus, an equilibrium is achieved, governed by the above equilibrium conditions. Indeed, conditions (15) are simply a restatement of Wardrop's (1952) first principle mathematically and mean that only those paths connecting an O/D pair will be used which have equal and minimal user travel costs. Otherwise, a traveler could improve upon his situation by switching to a path with lower travel cost.

Optimization Reformulation in a Special Case

In order to obtain a solution to the this problem, Beckmann, McGuire, and Winsten (1956) established that the solution to the equilibrium problem, in the case of separable user link travel cost functions, could be obtained by solving the following optimization problem:

$$\text{Minimize} \quad \sum_{a \in L} \int_0^{f_a} c_a(x)dx \tag{16}$$

subject to:

$$\sum_{p \in P_w} x_p = d_w, \quad \forall w \in W, \tag{17}$$

$$f_a = \sum_{p \in P} x_p \delta_{ap}, \quad \forall a \in L, \tag{18}$$

$$x_p \geq 0, \quad \forall p \in P. \tag{19}$$

Note that the conservation of flow equations are identical in both the user-optimized network problem (see (17)–(19)) and the system-optimized problem (see (7) – (9)). The behavior of the travelers, however, is different.

The objective function given by (16) is simply a device constructed to obtain a solution using general purpose convex programming algorithms. It does not possess the economic meaning of the objective function encountered in the system-optimization problem given by (6), equivalently, by (12).

Remark 1

A simplified user function, used in practice, sometimes also referred to as a common link performance function (see Sheffi 1985), is the expression developed by the Bureau of Public Roads (BPR). This equation is given by

$$c_a = c_a^0 \left[1 + \alpha (\frac{f_a}{t'_a})^{\beta} \right],$$

where, in this formula, c_a and f_a are the travel time and link load, respectively, on link a, c_a^0 is the free-flow travel time, and t'_a is the "practical capacity" of link a. The quantities α and β are model parameters, for which the values $\alpha = 0.15$ minutes and $\beta = 4$ are typical values. For example, these values imply that the practical capacity of a link is the flow at which the travel time is 15% greater than the free-flow travel time.

Toll Policies

I now describe how tolls, either in the form of path tolls or link tolls, can be imposed in order to make the system-optimizing solution also user-optimizing. Tolls serve as a means for modifying the travel cost as perceived by the individual travelers and are considered a powerful pricing policy instrument.

Recall that the system-optimizing flow pattern is one that minimizes the total travel cost over the entire network, whereas the user-optimized flow pattern has the property that no user has any incentive to make a unilateral decision to alter his/her travel path. One would expect the former pattern to be established when a central authority dictates the paths to be selected, so as to minimize the total cost in the system, and the latter, when travelers are free to select their routes of travel so as to minimize their individual travel cost. The latter solution, however, typically results in a higher total system cost and, in a sense, is an underutilization of the transportation network.

In order to remedy this situation tolls can be applied with the recognition that imposing tolls will not change the travel cost as perceived by society since tolls are not lost.

In particular, it will first be shown how tolls can be collected on a *link basis*, that is, every traveler on a link will be charged the same toll, irrespective of origin or final destination, or on a *path basis*, in which every traveler traveling from an origin to a destination on a particular path will be charged the same toll. (Note that one can construct multimodal versions of such a toll policy in which the pricing on links or paths is according to mode; see Dafermos 1973 and Nagurney 1999a).

Let t_a denote a toll associated with link a in the link-toll collection policy and let t_p denote the toll on path p associated with the path-toll collection policy. Of course, even in the link-toll collection policy one may define a "path toll" through the expression

$$t_p = \sum_{a \in L} t_a \delta_{ap}, \quad \forall p \in P. \tag{20}$$

Observe that, after the imposition of tolls, the travel cost as perceived by society remains $c_a(f_a)$, for all links $a \in L$. The travel cost on a path p as perceived by the individual, however, is modified to

$$\bar{C}_p = C_p(f) + t_p, \quad \forall p \in P. \tag{21}$$

Consequently, a system-optimizing flow pattern is still defined as before, that is, it is one that solves the problem

$$\text{Minimize}_{f \in K} \sum_{a \in L} \hat{c}_a(f_a), \tag{22}$$

where $K \equiv \{f |$ there exists an $x \geq 0$, satisfying (7) and (8)$\}$.

In particular, the solution to (22), under the assumption that each $\hat{c}_a(f_a)$ is convex, is equivalent to the conditions (13). Note that conditions (13), in turn, are equivalent to the statement For every O/D pair $w \in W$, there exists an ordering of the paths $p \in P_w$, such that

$$\hat{C}'_{p_1}(f) = \ldots = \hat{C}'_{p_{s_w}}(f) = \mu_w \leq \hat{C}'_{p_{s_{w+1}}}(f) \leq \ldots \leq \hat{C}'_{p_{n_w}}(f) \tag{23}$$

$$x_{p_{r_w}} > 0, \quad r_w = 1, \ldots, s_w$$

$$x_{p_{r_w}} = 0, \quad r_w = s_{w+1}, \ldots, n_w,$$

where n_w denotes the number of paths for O/D pair w.

On the other hand, in view of equilibrium conditions (15), one can deduce that the system-optimizing flow pattern x, after the imposition of a toll policy, is at the same time user-optimizing if: For every O/D pair $w \in W$, every path $p \in P_w$:

$$\bar{C}_{p_1}(f) = \ldots = \bar{C}_{p_{s_w}}(f) = \bar{\lambda}_w \leq \bar{C}_{p_{s_w+1}}(f) \leq \ldots \leq \bar{C}_{p_{n_w}}(f) \tag{24}$$

$$x_{p_{r_w}} > 0, \quad r_w = 1, \ldots, s_w$$

$$x_{p_{r_w}} = 0, \quad r_w = s_{w+1}, \ldots, n_w.$$

The following proposition (see Dafermos (1973)) enables the construction of toll policies which render an S-O flow pattern also a U-O one:

Proposition 1

A toll-collection policy renders a system-optimizing flow pattern user-optimizing if and only if for each O/D pair $w \in W$:

$$
\begin{aligned}
t_{p_1} &= \bar{\lambda}_w - C_{p_1}(f) \\
\vdots \quad & \qquad\quad \vdots
\end{aligned}
\tag{25}
$$

$$
\begin{aligned}
t_{p_{s_w}} &= \bar{\lambda}_w - C_{p_{s_w}}(f) \\
t_{p_{s_{w+1}}} &\geq \bar{\lambda}_w - C_{p_{s_{w+1}}}(f) \\
\vdots \quad & \qquad\quad \vdots \\
t_{p_{n_w}} &\geq \bar{\lambda}_w - C_{p_{n_w}}(f).
\end{aligned}
\tag{26}
$$

Proof:

It is clear that if (23) and (24) are satisfied for the same flow pattern x, then (25) and (26) follow. Conversely, if (25) and (26) are satisfied, then any f that satisfies (23) also satisfies (24).

I now turn to the determination of the link-toll and the path-toll collection policies.

Solution of the Link-Toll Collection Policy

Using (23), (24), (25), and (26), one reaches the conclusion that the link-toll collection policy is determined by

$$t_a = \frac{\partial \hat{c}_a(f_a)}{\partial f_a} - c_a(f_a), \quad \forall a \in L, \tag{27}$$

where both the first and the second terms on the right-hand side of expression (27) are evaluated at the system-optimizing solution f.

Hence, to determine the link-toll policy, one first must compute the system-optimizing solution. This can be accomplished using a general-purpose convex programming algorithm, an appropriate nonlinear network code, or, in the case of separable linear user cost functions, the equilibration algorithm of Dafermos and Sparrow (1969). Once the system-optimizing solution is established, one then substitutes the S-O load pattern f into equation (2.27) to compute the link toll t_a for all links $a \in L$.

Solution of the Path-Toll Collection Policy

It is clear from (25) and (26) that one may construct an infinite number of solutions of the path-toll collection problem. For example, one may select, a priori, for each $w \in W$, the level of personal travel cost $\bar{\lambda}_w$, as well as the values of $t_{p_{s_{w+1}}}, \ldots, t_{p_{n_w}}$, subject to only constraint (26), and then determine a path-toll pattern according to (25) and (26). Hence, in this case there is some flexibility in selecting a toll pattern, and one can incorporate additional objectives. Certain possibilities are:

(i) One may wish to ensure that some, if not all, travelers are charged with a nonnegative toll; in other words, no subsidization is allowed. This can be accomplished by choosing the corresponding $\bar{\lambda}_w$ sufficiently large.

(ii) Suppose one wishes a "fair" policy. A possible one would be to ensure that the level of personal travel cost $\bar{\lambda}_w$ is equal to the personal travel cost λ_w before the imposition of tolls.

In summary, one computes the path-toll policy as follows. First, compute the system-optimizing solution. Then determine the user travel cost C_p, for all paths $p \in P$, evaluated at the system-optimizing solution and $\bar{\lambda}_w$, $\forall w \in W$, so that an objective is met. Finally, compute the path tolls t_p, $\forall p \in P$, according to (25) and (26).

A simple example is now presented in order to illustrate how one computes a link-toll policy.

Example 1

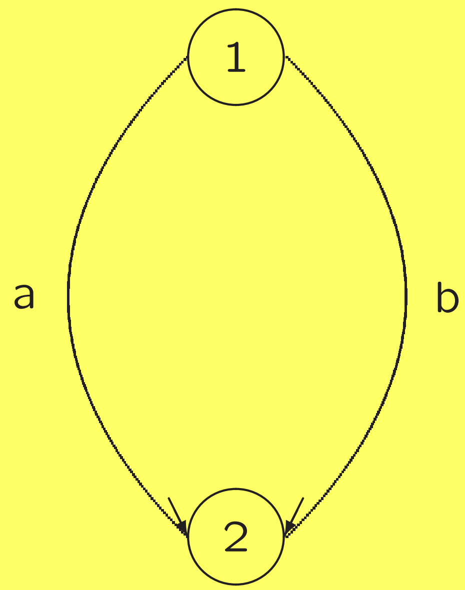
Consider the network depicted in the next Figure in which there are two nodes: 1, 2; two links: a, b; and a single O/D pair $w_1 = (1, 2)$. Let path $p_1 = a$ and path $p_2 = b$.

Assume, for simplicity, the user link travel cost functions:

$$c_a(f_a) = 2f_a + 5, \quad c_b(f_b) = f_b + 10,$$

and the travel demand:

$$d_{w_1} = 10.$$



Network topology for a link-toll policy example

In the absence of any policies, travelers operating in a user-optimized manner will select the paths as follows: $x_{p_1} = 5$, and $x_{p_2} = 5$ with induced link load patterns of: $f_a = 5$ and $f_b = 5$. The incurred user travel costs on the paths under this user-optimized flow pattern will be:

$$C_{p_1} = c_a = 15, \quad C_{p_2} = c_b = 15,$$

which satisfies the traffic equilibrium conditions (15).

This path flow pattern, in turn, will yield a total cost on the network given by $c_a \times f_a + c_b \times f_b = 75 + 75 = 150$.

The system-optimized flow pattern satisfying conditions (13) is, however, given by: $x_{p_1} = 4\frac{1}{6}$, $x_{p_2} = 5\frac{5}{6}$, which induces the link load pattern: $f_a = 4\frac{1}{6}$, $f_b = 5\frac{5}{6}$ and the marginals of the total travel costs on the paths are:

$$\hat{C}'_{p_1} = \hat{c}'_a = 21\frac{2}{3}, \quad \hat{C}'_{p_2} = \hat{c}'_b = 21\frac{2}{3},$$

with a total cost in the network under the S-O pattern equal to $131\frac{7}{18}$, which is clearly lower than the total cost under the U-O flow pattern above, which was 150.

I now turn to the computation of the link-toll policy (cf. (27)) which will render the S-O flow pattern also a U-O flow pattern. The link-toll policy that renders the system-optimizing flow pattern also user-optimized is given by:

$$t_a = 8\frac{1}{3}, \quad t_b = 5\frac{5}{6},$$

with the induced user costs (cf. (21)) $\bar{C}_{p_1} = \bar{C}_{p_2} = 21\frac{2}{3}$.

Models with Asymmetric Link Costs

The past several decades have been witness to much dynamic research activity in both the modeling and the development of methodologies to enable the formulation and computation of more general traffic network equilibrium models.

Examples of general models include those that allow for multiple modes of transportation or multiple classes of users, who perceive cost on a link in an individual way.

In this part of the lecture, I consider traffic network models in which the user travel cost on a link is no longer dependent solely on the flow on that link.

Other traffic network models, including dynamic traffic models, can be found in Mahmassani et al. (1993), and in the books by Ran and Boyce (1996) and Nagurney and Zhang (1996), and the references therein.

Link- and Path-Toll Policies

I now consider user link travel cost functions which are of the general form (refer to (3)), where the travel cost on a link may depend also on the load of this as well as other loads on the network, that is,

$$c_a = c_a(f), \quad \forall a \in L. \tag{28}$$

In the case where the symmetry assumption exists, that is, $\frac{\partial c_a(f)}{\partial c_b(f)} = \frac{\partial c_b(f)}{\partial c_a(f)}$, for all links $a, b \in L$, one can still reformulate the solution to the traffic network equilibrium problem satisfying equilibrium conditions (15) as the solution to an optimization problem (cf. Nagurney 1999a and the references therein), albeit, again, with an objective function that is artificial and simply a mathematical device.

However, when the symmetry assumption is no longer satisfied, such an optimization reformulation no longer exists and one must appeal to *variational inequality theory*.

Indeed, it was in the problem domain of traffic network equilibrium problems that the theory of finite-dimensional variational inequalities realized its earliest success, beginning with the contributions of Smith (1979) and Dafermos (1980).

For an introduction to the subject, as well as applications ranging from traffic network equilibrium problems to financial equilibrium problems, see the book by Nagurney (1999a). The methodology of finite-dimensional variational inequalities is utilized in this book in order to develop a spectrum of policy models.

The system-optimization problem, in turn, in the case of nonseparable user link travel cost functions becomes (see also (6)–(9)):

$$\text{Minimize} \quad \sum_{a \in L} \hat{c}_a(f), \tag{29}$$

subject to (7)–(9), where $\hat{c}_a(f) = c_a(f) \times f_a$, $\forall a \in L$.

The system-optimality conditions remain as in (13) but where now the marginal of the total cost on a path becomes, in the more general case:

$$\hat{C}'_p = \sum_{a,b \in L} \frac{\partial \hat{c}_b(f)}{\partial f_a} \delta_{ap}, \quad \forall p \in P. \tag{30}$$

The link-toll collection policy (see (27)) is now given by:

$$t_a = \sum_{b \in L} \frac{\partial \hat{c}_b(f)}{\partial f_a} - c_a(f), \quad \forall a \in L. \tag{31}$$

The path-toll policy, in turn, can still be computed according to formulas (25) and (26), where $\bar{C}_p$ is given by (21).

See also Bergendorff, Hearn, and Ramana (1997) for other toll policies for traffic networks with asymmetric user link travel cost functions.

Practical Considerations

Tolls, hence, provide an example of road pricing for congestion management. As noted by Small and Gomez-Ibanez (1998), public officials have become increasingly interested in congestion pricing as a means to stem the growth of traffic congestion. Refer to the Table for various road pricing schemes in practice.

Practical experience with road pricing has been increasing globally, with notable examples including those of Singapore, Hong Kong, and Cambridge, England.

Singapore's area license scheme was initiated in 1975 and is still operational. It has recently adopted an electronic version. Hong Kong has an electronic road pricing scheme, whereas Cambridge has investigated congestion pricing schemes which have included also rigorous modeling and theoretical underpinnings.

Scandinavian cities, in turn, have adopted a type of road pricing which, although the systems do not represent congestion pricing, do support highway financing. Specifically, toll rings now surround the three Norwegian cities of Oslo, Trondheim, and Bergen and such a ring is also in the planning stages for Stockholm.

Various road pricing schemes in practice

Type of road pricing	In place	Under study
City center: congestion pricing	Singapore, 1975	Hong Kong Cambridge, UK
City center: tolling	Bergen, 1986 Oslo, 1990 Trondheim, 1991	Stockholm
Single facility: congestion pricing	Autoroute A1, France, 1992 Route R1, California, 1995 Interstate 15, San Diego, 1996	
Area-wide: congestion pricing		Ramstad London

Source: Small and Gomez-Ibanez (1998)

France on Autoroute A1, which is an expressway connecting Paris to Lille, instituted congestion pricing, as did California in Riverside County.

Indeed, the first site of congestion pricing in the United States is a section of highway in southern California which opened in 1995.

During the past decade, interest has grown both in the United States and abroad, especially in Europe, to utilize advanced computer, electronics, and communication technologies (collectively referred to as intelligent transportation systems or ITS) in order to improve transportation efficiency (as well as safety). With the advent of such associated developments as electronic toll technology, as well as intelligent vehicle highway systems (IVHS) and advanced traffic information management systems (ATIMs) (cf. Boyce, Kirson, and Schofer 1994), it is becoming increasingly pragmatic to implement both the tolls described in this chapter, as well as other road pricing and and permit schemes that are described in this book and which focus both on congestion reduction and emission abatement.

Variational Inequality Formulations of Fixed Demand Problems

As mentioned earlier, in the case where the user link travel cost functions are no longer symmetric, one cannot compute the solution to the U-O, that is the traffic network equilibrium, problem using standard optimization algorithms. Such cost functions are very important from an application standpoint since they allow for asymmetric interactions on the network.

For example, allowing for asymmetric cost functions permits one to handle the situation when the flow on a particular link affects the cost on another link in a different way than the cost on the particular link is affected by the flow on the other link.

Returning to the network depicted in the first Figure, an example of an asymmetric user link travel cost structure would be:

$$c_a(f) = 5f_a + f_b + 10, \quad c_b(f) = 3f_b + 2f_a + 15,$$

since $\frac{\partial c_a}{\partial f_b} = 1 \neq \frac{\partial c_b}{\partial f_a} = 2$.

The allowance for such asymmetric interactions enables also the more realistic modeling of multimodal traffic networks in which a particular flow of a mode affects the costs of other modes in a different manner than it is affected by the other modes.

Since in this course equilibrium is such a fundamental concept in terms of sustainable transportation networks and since variational inequality theory is one of the basic ways in which to study such problems I now, for completeness, also give variational inequality formulations of the traffic network equilibrium conditions (15).

These formulations are presented without proof (for derivations, see Smith 1979 and Dafermos 1980, as well as Florian and Hearn 1995 and the book by Nagurney 1999a).

Moreover, appropriate variational inequalities for the specific models are derived later in this course.

First, the definition of a variational inequality problem is recalled. I then give both the path flow formulation as well as the link load formulation for the traffic network equilibrium conditions.

Specifically, the variational inequality problem (finite-dimensional) is defined as follows:

Definition 1 (Variational Inequality Problem)

The finite-dimensional variational inequality problem, VI$(F, \mathcal{K})$*, is to determine a vector* $X^* \in \mathcal{K}$ *such that*

$$\langle F(X^*)^T, X - X^* \rangle \geq 0, \quad \forall X \in \mathcal{K}, \tag{32}$$

where F *is a given continuous function from* $\mathcal{K}$ *to* R^N, $\mathcal{K}$ *is a given closed convex set, and* $\langle \cdot, \cdot \rangle$ *denotes the inner product in* R^N*.*

Variational inequality (32) is referred to as being in *standard form*. Hence, for a given problem, typically an *equilibrium* problem, one must determine the function F that enters the variational inequality problem, the vector of variables X, as well as the feasible set $\mathcal{K}$.

The variational inequality problem contains, as special cases, such well-known problems as systems of equations, optimization problems, and complementarity problems.

Hence, it is a powerful unifying methodology for equilibrium analysis and computation.

Theorem 1 (Variational Inequality Formulation of Traffic Network Equilibrium with Fixed Demands – Path Flow Version)

A vector $x^* \in K^1$ is a traffic equilibrium path flow pattern, that is, it satisfies equilibrium conditions (15) if and only if it satisfies the variational inequality problem:

$$\sum_{w \in W} \sum_{p \in P_w} C_p(x^*) \times (x - x^*) \geq 0, \quad \forall x \in K^1, \tag{33}$$

or, in vector form:

$$\langle C(x^*)^T, x - x^* \rangle \geq 0, \quad \forall x \in K^1, \tag{34}$$

where C is the n_P-dimensional column vector of path user travel costs and K^1 is defined as:

$K^1 \equiv \{x \geq 0$, *such that* (17) *holds*$\}$.

Theorem 2 (Variational Inequality Formulation of Traffic Network Equilibrium with Fixed Demands – Link Load Version)

A vector $f^* \in K^2$ is a traffic equilibrium link load pattern if and only if it satisfies the variational inequality problem:

$$\sum_{a \in L} c_a(f^*) \times (f_a - f_a^*) \geq 0, \quad \forall f \in K^2, \tag{35}$$

or, in vector form:

$$\langle c(f^*)^T, f - f^* \rangle \geq 0, \quad \forall f \in K^2, \tag{36}$$

where c is the n-dimensional column vector of link user travel costs and K^2 is defined as:
$K^2 \equiv \{f \,|\, \textit{there exists an}\, x \geq 0$
and satisfying (17) *and* (18)$\}$.

Note that one may put variational inequality (34) in standard form (32) by letting $F \equiv C$, $X \equiv x$, and $\mathcal{K} \equiv K^1$. Also, one may put variational inequality (36) in standard form where now $F \equiv c$, $X \equiv f$, and $\mathcal{K} \equiv K^2$.

Variational Inequality Formulations of Elastic Demand Problems

Now, the general traffic network equilibrium model with elastic travel demands due to Dafermos (1982) is recalled. Specifically, it is assumed that now one has associated with each O/D pair w in the transportation network a travel disutility λ_w, where here the general case is considered in which the travel disutility may depend upon the entire vector of travel demands, which are no longer fixed, but are now variables, that is,

$$\lambda_w = \lambda_w(d), \quad \forall w \in W, \tag{37}$$

where d is the J-dimensional column vector of the travel demands.

Conservation of Flow Equations for Elastic Demand Model

The notation, otherwise, is as described earlier, except that here I also consider user link travel cost functions which are general, that is, of the form (28). The conservation of flow equations (see also (1) and (2)), in turn, are given by

$$f_a = \sum_{p \in P} x_p \delta_{ap}, \quad \forall a \in L, \tag{38}$$

$$d_w = \sum_{p \in P_w} x_p, \quad \forall w \in W, \tag{39}$$

$$x_p \geq 0, \quad \forall p \in P. \tag{40}$$

Hence, in the elastic travel demand case, the travel demands in expression (39) are now variables and no longer given, as was the case for the fixed travel demand expression in (2).

Traffic Network Equilibrium Conditions in the Case of Elastic Travel Demand

The traffic network equilibrium conditions (see also (15)) now take on in the elastic travel demand case the following form: For every O/D pair $w \in W$, and each path $p \in P_w$, a vector of path flows and travel demands (x^*, d^*) satisfying (39)–(40) (which induces a link load pattern f^* through (38)) is a traffic network equilibrium pattern if it satisfies:

$$C_p(x^*) \begin{cases} = \lambda_w(d^*), & \text{if} \quad x_p^* > 0 \\ \geq \lambda_w(d^*), & \text{if} \quad x_p^* = 0. \end{cases} \tag{41}$$

Equilibrium conditions (41) state that the travel costs on used paths for each O/D pair are equal and minimal and equal to the travel disutility associated with that O/D pair. Travel costs on unutilized paths can exceed the travel disutility.

In the next two theorems, both the path flow version and the link load version of the variational inequality formulations of the traffic network equilibrium conditions (41) are presented. These are analogues of the formulations (33) and (34), and (35) and (36), respectively, for the fixed demand model.

Theorem 3 (Variational Inequality Formulation of Traffic Network Equilibrium with Elastic Demands – Path Flow Version)

A vector $(x^*, d^*) \in K^3$ *is a traffic equilibrium path flow pattern, that is, it satisfies equilibrium conditions (41) if and only if it satisfies the variational inequality problem:*

$$\sum_{w \in W} \sum_{p \in P_w} C_p(x^*) \times (x - x^*) - \sum_{w \in W} \lambda_w(d^*) \times (d_w - d_w^*) \geq 0,$$

$$\forall (x, d) \in K^3, \tag{42}$$

or, in vector form:

$$\langle C(x^*)^T, x - x^* \rangle - \langle \lambda(d^*)^T, d - d^* \rangle \geq 0, \quad \forall (x, d) \in K^3, \tag{43}$$

where λ *is the* J*-dimensional vector of travel disutilities and* K^3 *is defined as:* $K^3 \equiv \{x \geq 0, \textit{ such that } (39) \textit{ holds}\}$.

Theorem 4 (Variational Inequality Formulation of Traffic Network Equilibrium with Elastic Demands – Link Load Version)

A vector $(f^*, d^*) \in K^4$ *is a traffic equilibrium link load pattern if and only if it satisfies the variational inequality problem:*

$$\sum_{a \in L} c_a(f^*) \times (f_a - f_a^*) - \sum_{w \in W} \lambda_w(d^*) \times (d_w - d_w^*) \geq 0,$$

$$\forall (f, d) \in K^4, \tag{44}$$

or, in vector form:

$$\langle c(f^*)^T, f - f^* \rangle - \langle \lambda(d^*)^T, d - d^* \rangle \geq 0, \quad \forall (f, d) \in K^4, \tag{45}$$

where $K^4 \equiv \{(f, d)$, *such that there exists an* $x \geq 0$ *satisfying* (38), (39)$\}$

Note that, under a symmetry assumption on the travel disutility functions, in addition to such an assumption on the user link travel cost functions, one can obtain (see Beckmann, McGuire, and Winsten 1956) an optimization reformulation of the traffic network equilibrium conditions (41), which in the case of separable user link cost functions and travel disutility functions takes is given by:

$$\text{Minimize} \quad \sum_{a \in L} \int_0^{f_a} c_a(x)dx - \sum_{w \in W} \int_0^{d_w} \lambda_w(y)dy \qquad (46)$$

subject to: (38)–(40).

I now present an example of an elastic demand traffic network equilibrium problem.

Example 2

Consider the network depicted in the next Figure in which there are three nodes: 1, 2, 3; three links: a, b, c; and a single O/D pair $w_1 = (1, 3)$. Let path $p_1 = (a, c)$ and path $p_2 = (b, c)$.

Assume that the user link travel cost functions are:

$$c_a(f) = 5f_a + 2f_b + 5, \; c_b(f) = 7f_b + f_a + 5,$$

$$c_c(f) = 3f_c + f_a + f_b + 7,$$

and the travel disutility is: $\lambda_{w_1}(d_{w_1}) = -2d_{w_1} + 99$.

The U-O flow and demand pattern that satisfies equilibrium conditions (41) is:

$$x^*_{p_1} = 5, \quad x^*_{p_2} = 4,$$

$$d^*_{w_1} = 9,$$

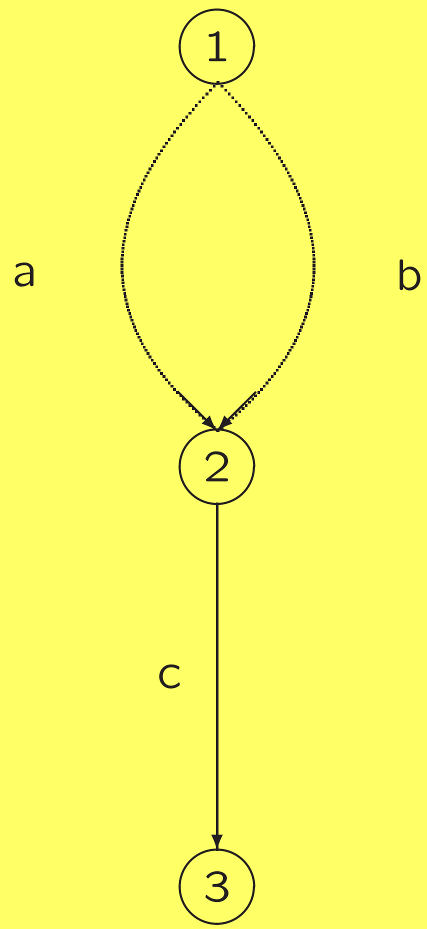
with associated link load pattern:

$$f^*_a = 5, \quad f^*_b = 4, \quad f^*_c = 9.$$

The incurred user travel costs on the paths are:

$$C_{p_1} = C_{p_2} = 81,$$

which is precisely the value of the travel disutility λ_{w_1}. Hence, this flow and demand pattern satisfies equilibrium conditions (41).



An elastic demand example