Efficient Transportation in Networks

Martin Beckmann

November 7, 1952

1.1 Introduction The regular occurrence of congestion in our highway system, so conspicuous a source of economic loss, not only raises problems to the traffic authorities concerned with handling an increasing demand for the use of roads, but also points to the existence of an important allocation problem for available road capacity in the context of any large undertaking that involves highway transportation on a considerable scale. The purpose of this paper is to inquire into the nature of the economic relationships that underly the phenomenon of congestion, to develop a model in terms of which the data gathered by traffic engineering sources on highway capacity may be brought to bear on this allocation problem, and to develop criteria of efficient allocation of road capacity which can be applied to the actual computation of the problem solution.

To this end we consider the following theoretical problem which is a development from earlier studies in activity analysis of transportation [1].

(1) Given a network of transportation channels whose capacities are given functions of speeds, (2) given also the capacities of intersections, independent of speeds, (3) given finally a transportation program prescribing net amounts
of commodities to be shipped from or received at given terminal points; what combination of traffic flows and speeds minimizes the inputs of transportation resources required?

As a by-product of the solution to this problem one may expect to find the social cost of transportation in a congested network, the determination of which is theoretically interesting because of its implications for efficient location of economic activities.

Finally the analysis of the allocation problem may be expected to lead to an economic evaluation of the network capacities in terms of the short run costs of congestion engendered by the loss of each unit of given road or intersection capacity.

1.2. Economic Plausibility Considerations. The phenomenon of congestion may be approached in two steps. Assume first that speeds are fixed on all roads or that capacity does not depend significantly on speed. In that case congestion is present whenever the demand of traffic for the use of a given road exceeds its (single-valued) capacity. Putting it this way shows at once the economic reasons for congestion. Roads tending to be congested are scarce resources rather than free goods for the public. Since there is however no market price to restrict the demand for this good, a mechanism of waiting queues must in the absence of direct measures of control reduce the demand to available capacity.

In the second place let speeds be variable and the capacities of a road be a function of (average) speed, having a maximum for some intermediate value of (average) speed. The notion of congestion arises with respect to the speed preferences of traffic. Congestion is present if the volume of traffic reduces the average speed thereby preventing the attainment of the preferred speed distribution. In the absence of direct allocation or of prices which equilibrate demand and capacity, the process of speed reduction expands
the supply, and the building up of waiting queues disperses demand until a correspondence is reached.

The break-down of normal flow conditions in traffic, which is what congestion amounts to, thus goes back in both cases to a failure of outside mechanisms to work toward an equilibrium. Of these mechanisms the two most obvious ones are those using allocation by direct control and those employing market prices. According to the theory of welfare economics, the latter procedure is, generally speaking, more efficient. Let us consider briefly how the operation of a price mechanism would affect the traffic equilibrium.

In the case of fixed capacities, a system of road tolls can presumably cut down the demand for each road to capacity and thereby eliminate both queues and congestion, provided of course that the transportation program is altogether possible within the limitations of the given network. Under speed variability the results of a balanced road toll system are less obvious. Generally speaking, congestion will not altogether disappear, since some degree of congestion on a road in high demand may be economical as compared with the cost of transportation over alternatives. But queues will be absent and traffic conditions more balanced. Considering the resulting speeds given and fixed each vehicle may choose its preferred route, and capacity at this speed will exactly permit the resulting traffic.

We may look at the tolls from still a different angle. In the case of fixed speed, a vehicle on a congested road causes a certain opportunity cost to traffic. This is the highest of the additional costs that must be born by vehicles excluded from the use of this road. In equilibrium the toll is just equal to this social cost. Under conditions of variable speed the social cost entailed by a vehicle on a congested road equals the cost to all other vehicles on the road of the delay caused by this vehicle. In equilibrium speeds also...
Those considerations cannot tell us, however, whether the marginal cost principle is sufficient to ensure an optimum allocation of resources and how to determine the equilibrium system of prices and flows. The substitution relations between flows of traffic and the interdependence of speeds, capacities and costs do not appear so simple as to permit definite conclusions on the strength of economic intuition.

1.3 Plan of this Paper  This paper attempts to buttress our intuitive arguments with an activity analysis of a network model of transportation. The objectives are: 1) to study the efficiency problem in terms of the minimization of a vector of physical inputs rather than money cost. 2) to establish that the marginal cost conditions of equilibrium are sufficient for efficiency in a broad class of cases. 3) to identify the present problem with a problem of saddle values for which computational procedures are well developed.

As a side objective it is hoped to obtain some information as to possibilities and limitations of the little explored field of nonlinear activity analysis.

The remaining parts of this section introduce the notation, discuss the assumptions and state theorems of activity analysis which are used later on. In the second section we take up our problem in a simplified version, considering speeds as given and fixed. The solution is characterized in terms of efficiency prices (Lagrangean parameters) and the efficiency conditions are seen to be associated with the saddle points of a function, which is further studied in section 3. The third section, concerned with the general case of variable speed, derives a condition for optimal speed. It is shown that this equation and the conditions obtained in section 2 are necessary and sufficient for the problem solution. There follows a brief account of the effect on transportation inputs of changes in intersection capacity and in the programs. In section 4 computational suggestions are made. The
concluding section summarizes the main results in terms of an allocation game.

1.4. Definitions and Assumptions. The key concepts in our problem are those of network, flow, capacity, transportation program, and transportation input. Presently we shall define these and state our assumptions concerning them.

1.4.1. By network we mean a connected system of traffic channels, of their junctions and of origination and destination points of traffic (Fig. 1).

For the sake of convenience we shall speak of roads, road-intersections and terminals respectively, intersections and terminals being denoted by indices i, roads by the (ordered) double indices ij referring to the adjacent locations.

Figure 1
Section from the highway map of New York State

(Four points are denoted by g, h, i and j respectively, and their connecting roads are marked by ij, gh, etc.)
We adopt the convention that any sums of terms with double indices \( i, j \) shall extend only over roads contained in the network.

1.4.2. A basic notion is that of commodity flow: the amount of commodity \( m \) that leaves point \( i \) per unit of time on vehicles entering road \( ij \), \( f_{ij}^{m} \).

Denote by \( \psi_{ij} \) the vector whose \( m \)th component is \( f_{ij}^{m} \)

\[
\psi_{ij} = (f_{ij}^{m})
\]

and by \( \varphi \) the vector composed of the \( \psi_{ij} \). The concept of flow makes possible the formulation of a discrete problem (optimum choice among a discrete number of route combinations) in terms of the minimization of a differentiable function of the continuous flows, subject, of course, to the obvious constraints of flow continuity in the network (cf. (1.5) infra). With the usual convention that if a sign such as "\( \varphi \)" holds for a pair of vectors then it holds for their respective components, we then have by definition of flows

\[
(1.1) \quad \varphi \geq 0
\]

We define traffic flow \( f_{ij} \) as the number of vehicles entering the road per unit time. If we exclude indivisibilities of equipment from this analysis, the flow of each commodity is strictly proportional to that of equipment by which this commodity is carried. By a proper choice of commodity units, traffic flow can be made equal to aggregate commodity flow.

\[
f_{ij} = \varepsilon^{'} \psi_{ij}
\]

where \( \varepsilon^{'} \) denotes the vector whose components are all 1.

1.4.3. The notion of vehicle speed \( u_{ij} \) is obvious. We assume uniform speed for all vehicles on a given road so as to exclude the problems involved in the passing of vehicles. While the latter constitute an important aspect essentially of traffic as such, it is believed that they do not change the character of the relationships between speed and capacity, a concave capacity function with respect to "average speed" replacing the differently shaped but also
concave capacity function defined for uniform speed (cf. below).

Denote the vector of $u_{ij}$ by $\mathbf{v} = (u_{ij})$.

(1.2) $\mathbf{v} \geq 0$

1.4.4 For every given speed $u_{ij}$, there exists a limit to the traffic flow $f_{ij}$ to be called the road capacity function $c_{ij}(u_{ij})$.

(1.3) $f_{ij} \leq c_{ij}(u_{ij})$.

Independently of the road capacities there may exist effective limits on the traffic through intersections. This calls for a notion of intersection capacity $c_i$. As a first approximation, intersection capacity may be considered a limit on aggregate flow through the intersection in which speed, being fixed at a safety level, does not enter.

(1.4) $\sum_j (f_{ij} + f_{ji}) \leq c_i$.

A detailed analysis of intersection capacity is possible only with reference to the individual layout and signalling apparatus of road intersections, a task well beyond the scope of this paper.

From traffic engineering data [2] $c_{ij}(u_{ij})$ is known to be a concave function with a finite maximum having approximately the shape indicated in figure 2. We assume it to have a continuous derivative $\frac{dc_{ij}}{du_{ij}}$. 

![Figure 2](image-url)
1.4.5 The quantities of commodity originating and terminating at each location are specified by the transportation program \( \gamma_i = (e_i^m) \).

\( e_i^m \) denotes the excess of receipts over shipments of commodity \( m \) at location \( i \). At all road intersections \( i \) which are not terminals \( \gamma_i = 0 \).

Now the excess of flows to location \( i \)
\[ \sum_j \psi_{ji} \]
over flows from location \( i \)
\[ \sum_j \psi_{ij} \]
must, under the static conditions which we assume, equal the net receipts \( \gamma_i \) at \( i \). That is, the transportation program requires

\[ (1.5) \quad \sum_j (\psi_{ji} - \psi_{ij}) = \gamma_i \]

Unless it is specified that each vehicle returns directly to its origin, there arises a problem of routing empty equipment. The excess availability of empty vehicles is found at each location to be
\[ \sum_{m=1}^{M} (x_{j1}^m - x_{ij}^m) = \sum_{m=1}^{M} e_{ij}^m. \]

Let this availability of empty equipment be denoted by \( -e_i^O \). The \( e_i^O \) may be considered as defining a residual program of empty shipping. The routing problem of empties is of course not different from that of any other commodity. Hence the flows of the commodity "empty equipment" may be considered included among the components of \( \psi \). With this convention

\[ (1.6) \quad \sum_i \gamma_i = 0 \]

for all \( i \).

We may assume the transportation program to be either fixed or a given function of the commodity prices at the location. It will be apparent in the course of the analysis that the two cases do not necessitate a separate treatment.
The transportation network is a closed system if we exclude exports and imports of transportation equipment and express exports and imports of commodities in terms of the programs of boundary locations. With this convention, the program of each commodity must balance

\[ \sum_{i} y_i = 0. \]

Note that (1.6) and (1.7) are restrictions on the data rather than the variables, flow and speed.

The constraints imposed on the traffic by the capacity limits of the network must be compatible with the transportation program, that is to say, there must exist at least one combination of flows and speeds which sustains the program within the limits set. In other words: we have to postulate the existence of some set of \( \psi, \psi \), such that the conditions (1.1)...(1.5) are all satisfied.

1.4.6. By transportation inputs we mean the mobile resources (as distinct from road capacity) employed in transportation. These include labor, rolling equipment (vehicles), fuel, and money costs representing other physical inputs. Our problem will be to minimize these inputs, assuming that these resources are available in unlimited quantities.

Since we do not take account here of indivisibilities the assumption is reasonable that, for a given transportation technique, all inputs into transportation on a given road and at a given speed are strictly proportional to flow. Since we are considering only one type of transportation at a time, say trucking, we may rule out the possibility of several technologies so that we have only one transportation activity on each road. Let

\[ k_{ij} = (k_{ij}^a) \]

represent the vector of resources used up by a unit of flow on road \( ij \). The total input associated with flows \( \psi \) is then
Now each input vector will be a function of speed \( v_{ij} \), a higher speed of transportation resulting in an increase in the use of some resources (fuel, money) and a decrease in the use of others (equipment, labor). In any case it is fairly safe to assume that inputs are convex functions of speed.\(^5\)

With respect to fuel traffic engineering data confirm the convex nature of this input function over the relevant range \([3]\). In the case of equipment and labor, convexity follows from the fact that these inputs are (approximately) proportional to transportation time, which is the reciprocal of speed, a clearly convex function. We assume the existence of continuous derivatives \( \frac{d k_{ij}}{d v_{ij}} \).

1.4.7. We are now in a position to state what is meant by an efficient combination of transportation activities. Efficiency is here defined in the sense of Pareto-optimality, the familiar concept that underlies the general theory of activity analysis. A combination of transportation activities is called efficient if it minimizes the input vector \( \mu \) subject to the constraints on \( \psi, \lambda \) given by the program and by the network capacities. A vector \( \mu \) is said to be minimal if there is no other vector \( \bar{\mu} \) satisfying the same constraints and such that \( \mu \preceq \bar{\mu} \), but not \( \mu = \bar{\mu} \).

In general, the ratios of resource inputs at a given speed are not the same for different roads. Hence substitution among the inputs is possible in a substitution among routes of transportation. The analysis will show however, that each efficient combination is characterized by a set of efficiency price ratios between the various resources. These may be interpreted as (inverse) ratios of convertibility of one resource to another so that we are again back to the case of one input. It is because of this
intrinsic convertibility that there exists a well defined notion of transportation cost also in this analysis based on input vectors.

1.5. The notation introduced in this section is summarized below; vectors are denoted by small Greek letters, scalars by small Roman letters.

Indices

\[ i \] locations

\[ i,j \] roads

\[ m \] commodities

\[ n \] resources

Variables

\[ \varphi = (\varphi_{ij}), \varphi_{ij} = (f_{ij}^m) \]

\[ f_{ij} = \delta^i \varphi_{ij} \]

\[ v = (u_{ij}) \]

\[ v = (u_{ij}) \]

\[ k_{ij} = (k_{ij}^n) \]

\[ \mu = \Sigma k_{ij} f_{ij} \]

\[ c_{ij}, c_{ij} \]

\[ g_{i}^m \]

The problem can now be stated mathematically as

Problem 1.

Find \( \varphi, v \) such that

\[ \mu (\varphi, v) = \min_{\varphi, v} \mu (\varphi, v) \]

subject to the constraints

(1.1) \[ \varphi \geq 0 \]

(1.2) \[ v \geq 0 \]

(1.3) \[ \delta^i \varphi_{ij} = c_{ij}^i (u_{ij}) \]

(1.4) \[ \Sigma_j \delta^i (\varphi_{ij} + \varphi_{ji}) \leq c_i \]

(1.5) \[ \Sigma_j (\varphi_{ij} - \varphi_{ji}) + \varphi_{i} = 0 \]
and such that \( \overline{y}, \overline{v} \) satisfy the conditions (1.1)-(1.5).

1.6. **Tools.** Before we proceed to the details of the problem solution, we shall state the two principal theorems on which the analysis is based.

The first one which may be called the main theorem of linear activity analysis, is stated in [4, Theorem 5.4.1, (p. 82)] to be referred to here as Lemma 1: "A necessary and sufficient condition that an attainable point \( y \) be efficient according to Definition 5.2 [a more general version of the efficiency definition on p. 10, was given with special reference to transportation problems] is that there exists a vector \( p \), normal to the possible cone \( A \) in \( y \), which has positive components for all final commodities, non-negative components for all primary commodities whose availability limit is reached in \( y \), and zero components for all primary commodities whose availability limit is not reached in \( y \),

\[
p_{\text{fin}} > 0, \quad p_{\text{pri}} > 0, \quad p_{\text{pri}}^\prime = 0. \]

The second theorem is a summary of well known extremum conditions in the calculus.

**Lemma 2:** Let \( F(x, y) \) be differentiable, jointly convex in \( x \), and jointly concave in \( y \); \( x = (x_i) \), \( y = (y_j) \). Then \( F \) has a saddle value over the positive orthant of \( x, y \) precisely where

\[
\frac{\partial F}{\partial x_i} \begin{cases} > \end{cases} 0 \quad \text{if} \quad x_i \begin{cases} > \end{cases} 0
\]

\[
\frac{\partial F}{\partial y_j} \begin{cases} > \end{cases} 0 \quad \text{if} \quad y_j \begin{cases} > \end{cases} 0
\]

[cf. 5; Lemma 3 (p. 485)].

2.1. A Linear Problem. Suppose that speeds are given and fixed at arbitrary (not necessarily optimal) values. Then input and capacity functions are linear functions of flows and we obtain a linear problem. 8/

Problem 2.

Find all \( \varphi \) such that

\[
\sum_{i,j} \kappa_{ij}(\varepsilon \cdot \varphi_{ij}) = \text{Min} \sum_{i,j} \kappa_{ij}(\varepsilon \cdot \varphi_{ij})
\]

subject to the constraints

\[
\varphi \geq 0
\]

\[
\sum_j (\varphi_{ij} - \varphi_{ji}) \geq y_i = 0
\]

\[
\varepsilon' \varphi_{ij} \geq c_{ij}
\]

\[
\sum_j \varepsilon'(\varphi_{ij} + \varphi_{ji}) \geq c_i
\]

In treating extremum problems of linear vector functions subject to linear equalities and inequalities as constraints we are on familiar ground. However, Problem 2 is not in the form of a standard activity analysis problem. We shall now show that it can be reduced to the following one.

Problem 3.

activities

\( f_{ij} \)

shipments

inputs (primary commodities)

desired in themselves:

\( \sum_{ij} k_{ij} f_{ij} \)

mobile transportation resources

subject to availability limitations:

\( f_{ij} \leq c_{ij} \)

road capacities
intersection capacities

\[ \sum_{j} (\ell_{ij} + f_{ij}) \frac{1}{c_{i}} \]

net shipments from origination points

\[ r_{i} = \sum_{j} (r_{ij}^{m} - f_{ij}^{m}) \frac{1}{c_{i}} \]

where \( i, m \in \{ i, m : g_{i}^{m} \leq 0 \} \)

outputs (final commodities)

net receipts at destination points

\[ y_{i}^{m} = \sum_{j} (f_{ji}^{m} - f_{ij}^{m}) \]

where \((i, m) \in \{ i, m : g_{i}^{m} > 0 \} \).

It is intuitively clear and follows from the definitions just

given that the sum of flow outputs of each commodity must equal the sum

\[ \sum_{i: \text{destinations}} y_{i}^{m} = \sum_{i: \text{origins}} x_{i}^{m} = \sum_{i: \text{origins}} (f_{ji}^{m} - f_{ij}^{m}) = 0 \]

Hence

\[ \sum_{i: \text{destinations}} y_{i}^{m} = \sum_{i: \text{origins}} x_{i}^{m} = \sum_{i: \text{origins}} g_{i}^{m} \]

(cf. equation (1.5)). Therefore the point \( y_{i}^{m} = g_{i}^{m} \) in the "space of final commodities" is "efficient" (in the sense in which this term is used in

activity analysis [4, Def. 5.2., p. 79]), provided that the inputs of de-
sired commodities are minimized. This shows that the solutions of Problem

2 can be obtained as a subset of the efficient point set of activity analy-
sis Problem 3. This subset is characterized by the condition

\[ \sum_{j} (\varphi_{ij} - \varphi_{ji}) + y_{i} = 0 \quad \text{for all } i. \]

The verification that the postulates of linear activity analysis [4, pp.

48-55] are satisfied by Problem 3 is straightforward and is left to the
reader. For instance, postulates C and D are nothing but our previous as-
sumption on the possibility of the transportation program within the limits
set by the network (p. 9 ).
We are now able to apply to Problem 3 an important theorem of activity analysis [4, 5.4.1, p. 82, in conjunction with definitions 5.1, 5.2 of p. 79] characterizing the efficient point set.

**Theorem 1:** Necessary and sufficient for \( \psi_{ij} \) to be a solution of Problem 2 is the existence of a vector of nonnegative numbers

\[
p = (r_{ij}, \ldots, r_i, \ldots, r_i^m, \ldots)
\]

such that in addition to the constraints

1. \( \psi \geq 0 \)
2. \( \varepsilon^i \psi_{ij} \leq c_{ij} \)
3. \( \sum_j \varepsilon^j (\psi_{ij} + \psi_{j1}) \leq c_i \)
4. \( \sum_j (\psi_{ij} - \psi_{j1}) + \gamma_i = 0 \)

the following conditions are satisfied for some preassigned vectors of positive numbers \( \Pi = (p^n) \)

\[
\Pi^i k_{ij} + r_{ij} + r_i + r_j \left\{ \begin{array}{l} \geq \varepsilon^i \psi_{ij} \leq c_{ij} \ 
\end{array} \right. \quad r_j^m - r_i^m
\]

if \( r_{ij} \geq 0 \)

where

1. \( r_{ij} \geq 0 \) if \( \varepsilon^i \psi_{ij} \leq c_{ij} \)
2. \( r_j \geq 0 \) if \( \sum_j \varepsilon^j (\psi_{ij} - \psi_{j1}) \leq c_i \)
3. \( r_i^m \geq 0 \) if \( (i, m) \left\{ \varepsilon \right\} \{ i, m: \varepsilon^m_1 > 0 \} \)

2.3. **Efficiency Prices.** The equations (2.1)...(2.4) may be called the efficiency conditions. They supplement the problem constraints (1.1), (1.3)-(1.5) for a complete determination of the solution of Problem 2.
If we regard the new parameters as prices in the following way --

\[ p^n \]  unit cost of resource \( n \)

\[ r_i \]  toll at intersection \( i \)

\[ r_{ij} \]  toll on road \( ij \)

\[ r^m_i \]  price of commodity \( m \) at terminal \( i \)

then the efficiency conditions take on the following economic meaning reminiscent of the role of real prices and tolls.

1. Geographical price differences are less than or equal to the cost of transportation consisting of costs of resources used plus tolls. Although equation (2.1) expresses this only for locations adjacent to one road, it is true a fortiori for any pair of locations.

2. If transportation of a commodity actually takes place between two given points the price difference between the receiving and the shipping point is equal to minimum transportation cost. Costs of transportation are equal to that minimum on all alternative routes used.

3. Suppose that alternative routes are used by flows of commodity \( m \) from \( i \) to \( j \). Then \( r^m_j = r^m_i \) being the same for all routes it is clear that \( r_{ij} + r_i + r_j \) must compensate for larger values \( \sum_{gh} \pi^g \kappa_{gh} \) on the alternative routes \( R \).

Now \( r_{ij} \) is common to all routes containing road \( ij \) and \( r_i \) to all routes passing intersection \( i \). We may conclude therefore that \( r_{ij} \), \( r_i \) are shadow prices compensating for the benefit of using this road or intersection, respectively, instead of a detour.

4. Prices of desirable resources are never zero, but tolls are zero whenever demand falls short of capacity.

5. Transportation costs being positive under all circumstances, circular shipments are excluded since the geographical price difference would be zero. This does not exclude the possibility of closed paths in the flow graph.
But the directions of flow are such that if we traverse the circuit adding transportation costs for clockwise flows and subtracting transportation costs for counter clockwise flows the sum vanishes.

Circuits that satisfy this condition of consistency with the existence of \( r_i^m \), are termed "neutral" [1, p. 248].

6. The resource prices and tolls do not discriminate among commodities. In particular the commodities excluded from the use of a desirable road are those for which relatively cheap substitute routes exist.

7. If resource prices and tolls are given the flows of each commodity represent a solution of a Hitchcock–Koopmans transportation problem in that commodity [4, pp. 222–259] with \( \gamma_r r_{ij} + c_{ij} + r_i + r_j \) the cost of transportation.

![Figure 3](image)

2.4. An Example. Consider a simple network (Figure 3) in which two traffic flows, north-bound and northeast-bound compete for the bottleneck 2 4 (see tables). Case 4 is remarkable in that detouring northbound traffic forces the northeast traffic out of its nearest route.
\[ k_{ij} = k_{ji} \]

\[
\begin{array}{c|ccccc}
\hline
& 1 & 2 & 3 & 4 & 5 \\
\hline
1 & 1 & & & & \\
2 & & & 3 & 2 & \\
3 & & & & 2 & \\
4 & & & & & 1 \\
\hline
\end{array}
\]

\[ g_{ij}^m \]

\[
\begin{array}{c|ccccc}
\hline
& 1 & 2 & 3 & 4 & 5 \\
\hline
1 & -1 & 1 & & & \\
2 & -1 & 1 & & & \\
\hline
\end{array}
\]

\[ o_{ij} \]

Case 1: 
\[ o_{23} \geq 1 \]
\[ c_{24} \leq 1 \]
\[ c_{23} + c_{24} = 2 \]

Case 2: 
\[ o_{23} < 1 \]
\[ c_{23} + c_{24} \leq 2 \]

Case 3: 
\[ o_{23} < 1 \]
\[ 1 = c_{23} + c_{24} \leq 2 \]

Case 4: 
\[ o_{23} + o_{24} = 1 \]

\[ \psi_{ij} \]

Case 1

Case 2
Figure 3.

$\rho_i = (r_i^m)$

and

$k_{ij} = \pi^{i} k_{ij}$.

Lemma 3: The function

\[\phi(\varphi, \Pi, \eta) = \sum_{ij} k_{ij} \psi_{ij} + \sum_{ij} r_{ij} (\zeta \psi_{ij} - c_{ij}) \cdot\]
\[ + \sum_{i} \sum_{j} \varepsilon^{i} \left( \psi_{ij} - \varphi_{j1} \right) \]
\[ + \sum_{i} \rho_{j} \left( \psi_{ij} - \psi_{j1} \right) + \gamma \]

has a saddle point \(\overline{\varphi}, \overline{\rho}\) for every solution \(\overline{\psi}\) and associated parameter vector \(\overline{\rho}\) of Problem 2.

\[ \max_{\rho \geq 0} G(\overline{\varphi}, \rho, \gamma) = G(\overline{\varphi}, \overline{\rho}, \gamma) = \min_{\varphi \geq 0} G(\varphi, \overline{\rho}, \gamma). \]

Here the parameter \(\gamma\) in \(G(\varphi, \rho, \gamma)\) indicates that the right side of (2.5) still depends on the \(u_{ij}\). The proof of the lemma is straightforward using the extremum conditions for linear functions (Lemma 2, end of Section 1).

**Corollary:** Let \(\overline{\psi}, \overline{\rho}\) be a solution of Problem 2. Then

\[ G(\overline{\varphi}, \overline{\rho}, \gamma) = \mu(\overline{\psi}, \gamma) = \sum_{ij} k_{ij} \varepsilon^{i} \psi_{ij}. \]

### 3. Analysis of Flows and Speeds.

3.1. The Original Problem Reformulated. In this section we turn back to our original Problem 1. Lemma 3 and Corollary imply that Problem 2 can be written in the form

\[ \min_{\varphi \geq 0} \max_{\rho \geq 0} G(\varphi, \rho, \gamma). \]

Since every solution of Problem 2 satisfies automatically the constraints of Problem 1 and because of the Corollary of Lemma 5 one obtains the following unconstrained version of Problem 1.

**Problem 4.**

\[ \min_{\gamma \geq 0} \min_{\varphi \geq 0} \max_{\rho \geq 0} G(\varphi, \rho, \gamma) \]
Consider the functions $k_{ij}(u_{ij}) = \mathcal{T} K_{ij}(u_{ij})$ and $c_{ij}(u_{ij})$ for a given road $ij$ (Figure k).

Now of the two values of $u_{ij}$ for which the function $c_{ij}(u_{ij})$ takes on a given value only that one can minimize inputs for which $k_{ij}(u_{ij})$ is smaller. Because of the convexity of $k_{ij}(u_{ij})$ and the concavity of $c_{ij}(u_{ij})$
this means that \( u_{ij} \) cannot be optimal unless it is contained in the interval between the points \( u^*_{ij} \) and \( u^{**}_{ij} \) at which, respectively, \( c_{ij}(u_{ij}) \) is maximal and \( k_{ij}(u_{ij}) \) is minimal. (If the two extrema coincide, so that \( u_{ij} = u^*_{ij} = u^{**}_{ij} \) then this is the optimal speed.) Since on this interval \( c_{ij} \) is monotonic, it follows that efficient speed can be defined as a unique and continuous function of efficient flows \( f_{ij} \) in the following way (Figure 5):

\[
 u_{ij} = \begin{cases} 
 u^{**}_{ij} & \text{for } 0 \leq f_{ij} \leq c_{ij}(u^{**}_{ij}) \\
 \text{inverse of } c_{ij}(u_{ij}) & \text{for } c_{ij}(u^{**}_{ij}) \leq f_{ij} \leq c_{ij}(u^*_{ij}) 
\end{cases}
\]

Since \( c_{ij}(u_{ij}) \) was assumed continuously differentiable, so is \( u_{ij}(f_{ij}) \) except at the point \( f_{ij} = c_{ij}(u^{**}_{ij}) \).

Problem 4 may now be reformulated, after substituting \( u_{ij} = (u_{ij}(f_{ij})) \) and observing that \( f_{ij} = \phi^* \psi_{ij} \):

Find: \( \min \max \ G(\phi, \rho ; (u_{ij}(\phi^*, \psi_{ij}))) \).
In explicit form, leaving out the terms \( \sum_{ij} r_{ij} (\mathcal{E} \cdot \psi_{ij} - c_{ij}) \),
which now vanishes identically, from \( G \) (equation 2.5), we have

**Problem 6.**

Find \( \min \max H(\psi, \rho) \) where
\[
\psi \neq 0, \quad \rho \neq 0
\]

\[
(3.1) \quad H(\psi, \rho) = \sum_{ij} k_{ij} (u_{ij}(\mathcal{E} \cdot \psi_{ij})) \mathcal{E} \cdot \psi_{ij}
\]
\[
+ \sum_{i} r_{i} \left( \sum_{j} \mathcal{E} \cdot (\psi_{ij} + \psi_{ji}) - c_{i} \right)
\]
\[
+ \sum_{i} \rho_{i} \left( \sum_{j} (\psi_{ij} - \psi_{ji}) + \gamma_{i} \right)
\]

3.2. A Saddle Value Problem. Suppose for a moment (as we shall show later) that

\[
k_{ij} (u_{ij}(\mathcal{E} \cdot \psi_{ij})) \mathcal{E} \cdot \psi_{ij}
\]
is a convex function of \( \psi_{ij} \). \( H \) is then a convex function of \( \psi \) and a
(linear and hence concave) function of \( \rho \). In this case, necessary and
sufficient for \( \bar{\psi}, \bar{\rho} \) to be a saddle point of \( H(\psi, \rho) \) is that (cf. Lemma 2, p. 12)

\[
\text{grad} \psi \begin{bmatrix} \mathcal{E} \cdot \psi_{ij} \end{bmatrix} = 0 
\]

componentwise as

\[
\begin{bmatrix} \mathcal{E} \cdot \psi_{ij} \end{bmatrix} = 0
\]

\[
\text{grad} \rho \begin{bmatrix} \mathcal{E} \cdot \psi_{ij} \end{bmatrix} = 0
\]

componentwise as

\[
\begin{bmatrix} \mathcal{E} \cdot \psi_{ij} \end{bmatrix} = 0
\]

and \( \bar{r}_{i} \begin{bmatrix} \mathcal{E} \cdot \psi_{ij} \end{bmatrix} > 0 \), respectively.

The first set of the equations (3.2) reads explicitly

\[
(3.3) \quad k_{ij} + r_{ij} \frac{d k_{ij}}{u_{ij}} + \frac{d u_{ij}}{d r_{ij}} + r_{i} + r_{j} \begin{bmatrix} \mathcal{E} \cdot \psi_{ij} \end{bmatrix} \mathcal{E} \cdot \psi_{ij} - \bar{r}_{i} \begin{bmatrix} \mathcal{E} \cdot \psi_{ij} \end{bmatrix} = 0
\]

From the definition of the function \( u_{ij}(r_{ij}) \) it is seen that the second
term in (3.3) is equal to

\[
(3.4) \begin{cases}
0 & \text{if } 0 \leq r_{ij} \leq c_{ij}(u_{ij}^*) \\
\frac{dk_{ij}}{du_{ij}} \frac{dc_{ij}}{du_{ij}} & \text{if } f_{ij} > c_{ij}(u_{ij}^*)
\end{cases}
\]

Since \( \frac{dk_{ij}}{du_{ij}} \) and \( \frac{dc_{ij}}{du_{ij}} \) have the same sign (see beginning of this section),
the second term in (3.3) may be set equal to an efficiency price \( r_{ij}^* \)

\[
(3.5) r_{ij}^* \begin{cases}
0 & \text{if } f_{ij} \begin{cases}
\leq & c_{ij}(u_{ij}^*)
\end{cases}
\end{cases}
\]

The definition of \( r_{ij}^* \) is

\[
\frac{dk_{ij}}{du_{ij}} c_{ij} - r_{ij}^* \frac{dc_{ij}}{du_{ij}} = 0.
\]

Except for the convexity of \( k_{ij}(u_{ij}(f_{ij}))f_{ij} \) we have proved

Theorem 2: The vectors \( \mathbf{v} = (r_{ij}^m), \mathbf{v} = (u_{ij}) \) are a solution of Problem 1
if and only if they satisfy the following conditions with some vector
\( \mathbf{p} = (r_{ij}, \ldots, r_i, \ldots, r_i^m, \ldots) \), in addition to the constraints (1.1)-(1.5)

\[
(3.5) r_j^m - r_j \begin{cases}
\leq & k_{ij}(u_{ij}) + r_{ij} + r_j \begin{cases}
\leq & 0
\end{cases}
\end{cases}
\]

\[
(3.6) \frac{dk_{ij}}{du_{ij}} c_{ij}(u_{ij}) - r_{ij} \frac{dc_{ij}}{du_{ij}} = 0
\]

where

\[
(3.7) r_{ij} \begin{cases}
\leq & 0
\end{cases} \text{ if } f_{ij} \begin{cases}
\leq & c_{ij}(u_{ij})
\end{cases}
\]

\[
(3.8) r_i \begin{cases}
\leq & 0
\end{cases} \text{ if } \sum_{j} (f_{ij} + f_{ji}) \begin{cases}
\leq & c_i
\end{cases}
\]

Here we have written \( r_{ij} \) for \( r_{ij}^* \) since the conditions (3.5) and (3.7) correspond exactly to conditions (2.1) and (2.2) of Theorem 1, thus showing the identity of the two efficiency prices.

3.3. Assertions about Convex Functions. The verification of the convexity
of $k_{ij}(u_{ij}, f_{ij})$ is routine by means of the following statements which are easily checked.

1. The inverse of a nondecreasing concave function is nondecreasing and convex.

2. The inverse of a nonincreasing concave function is nonincreasing and concave.

3. A nondecreasing convex function of a nondecreasing convex function is nondecreasing and convex.

4. A nonincreasing convex function of a nonincreasing concave function is nondecreasing and convex.

5. The product of two nonnegative, nondecreasing, convex functions is nondecreasing and convex.

<table>
<thead>
<tr>
<th>Statement</th>
<th>with functions</th>
<th>proves</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2</td>
<td>$a_{ij}(u_{ij})$</td>
<td>$u_{ij}(f_{ij})$ is either nondecreasing and convex or nonincreasing and concave</td>
</tr>
<tr>
<td>3, 4</td>
<td>$k_{ij}(u_{ij}, f_{ij})$</td>
<td>$k_{ij}(u_{ij}, f_{ij})$ is nondecreasing and convex</td>
</tr>
<tr>
<td>5</td>
<td>$k_{ij}(u_{ij}(x_{ij}), f_{ij})$</td>
<td>$k_{ij}(u_{ij}(f_{ij}), f_{ij})$ is nondecreasing and convex</td>
</tr>
</tbody>
</table>

3.4. Variation of Intersection Capacity. The question of how inputs of mobile resources react to variations of road and intersection capacities is an important one not only from a theoretical point of view but also with respect to the problem of economic gains from road (or intersection) construction, or losses from corresponding destruction. Losses (and gains) are here defined as the price of the additional inputs of mobile resources necessary to compensate for diminished capacity so as to maintain the given program. Thus the net gain from a (positive) change $dq_1$ of intersection
capacity \( c_1 \) is defined to be

\[
\frac{d}{d^+ c_1} \sum_{gh} k_{gh} \left( u_{gh}(x_{gh}) \right) f_{gh} \, dc_1.
\]

In order to calculate this expression recourse is had to equation (2.5) and the corollary of Lemma 3.

\[
(3.9) \quad \frac{d\varphi}{d^+ c_1} = \frac{\mathcal{G}}{\mathcal{C}_1} \cdot \left( \text{grad } \varphi \right) \cdot \frac{d\psi}{d^+ c_1} + \left( \text{grad } \rho \right) \cdot \frac{d\rho}{d^+ c_1}
\]

and similarly for \( \frac{d\varrho}{d^+ c_1} \). This distinction between the two directional derivatives is necessary because \( \varphi \) and \( \psi \) depend on the parameters of the constraints in a piece-wise continuous fashion, as can be shown from the general theory of activity analysis. Since the values of \( \varphi \) and \( \psi \) at the jump discontinuities can be chosen so as to make \( \varphi(c_1) \) and \( \psi(c_1) \) continuous either from the right or from the left, there exist in these cases either right or left hand derivatives, and the chain rule of differentiation is valid.

We remark next that the composite terms in (3.9) must vanish. Consider, for instance, the member \( \frac{\mathcal{G}}{\mathcal{C}_1} \cdot \frac{d\varphi}{d^+ c_1} \). Now there exists an open right-hand neighborhood of \( c_1 \) in which either \( \varphi \) or \( \psi \) vanishes identically or is strictly positive. In the first case \( \frac{d\varphi}{d^+ c_1} = 0 \), in the second

\[
\frac{\mathcal{G}}{\mathcal{C}_1} = f_{gh} - c_{gh} = 0 \text{ by (2.2). Hence } \frac{\mathcal{G}}{\mathcal{C}_1} \cdot \frac{d\varphi}{d^+ c_1} = 0 \text{ in any case; similarly it follows that the rest of the composite derivatives vanish. Thus}
\]

\[
(3.10) \quad \frac{d\varphi}{d^+ c_1} = \frac{\mathcal{G}}{\mathcal{C}_1} = - p_1
\]

where the + subscript assigns that value to \( p_1(c_1) \) which renders it a function continuous from the right (cf. above). There may thus result two different rates of positive and negative substitution.
The question whether this marginal rate of substitution is nondecreasing, will not be settled here. In the case of fixed speeds, the answer is in the affirmative as can be seen from general principles of linear activity analysis [4, Section 4.10] or the (weak) "Le Chatelier principle" [7]. The case of variable speeds is not amenable to treatment in terms of the "Le Chatelier principle", and there is reason to suspect that the marginal rate of substitution is not always nondecreasing.

Whenever it is, however, then the term \(-p_1 \triangle c_1\) gives an estimate of the minimum loss to the system in terms of expenses for mobile resources due to the destruction of intersection capacity \(\triangle c_1\).

In the case of road capacity it is not possible to define a capacity loss unless it has been specified how the capacity function \(c_{ij}(u_{ij})\) depends on the physical dimensions of a road. This subject is in need of further clarification.

3.5. Variation of the Program, The Marginal Cost of Transportation. The same considerations as in the last section may be used to calculate the cost in terms of resource prices of a change in the program. Suppose that the transportation program of a commodity \(m\) calls for the shipment of \(g\) additional units from \(i\) and the receipt of \(g\) additional units at \(j\). Replacing \(\varepsilon^m_i, \varepsilon^m_j\) by \(\varepsilon^m_i - g, \varepsilon^m_j - g\), respectively, and differentiating the expression for cost, \(G(\varphi, \rho, \gamma)\) of equation (2.5), with respect to \(g\) we see that

\[
\frac{dG}{d_g} = \sum_i \frac{\partial G}{\partial g} = -p_i^m + p_j^m.
\]

With the same reservation as for two-valued rates of substitution in the two directions it can be concluded therefore that \(r_j^m = r_i^m\) satisfying

\[
(3.11) \quad r_j^m - r_i^m < k_{ij} + r_{ij} + r_i + r_j,
\]
is the marginal rate of substitution of resources, evaluated at preassigned
prices $p^n$, for programmed origination at point $i$ and termination at point $j$
of commodity $m$. In particular, if direct transportation of commodity $m$ be-
tween these terminals is taking place, the "*" sign holds in (3.11) and

$$k_{ij} + r_{ij} + r_i + r_j$$

is seen to equal the marginal cost of transportation on road $ij$.

The remarkable feature of this expression for marginal cost of trans-
portation is that in addition to the price $k_{ij}$ of inputs directly attribut-
able to this flow there arises a shadow toll, $r_{ij} + r_i + r_j$, indicating
that an opportunity cost is caused by the use of road and intersection ca-
pacities. To be sure, this (nonnegative) toll is strictly positive only
when the capacity limits in question are reached, that is, in the presence
of congestion.

To put the same fact differently, the term $k_{ij} = \Pi^r \kappa_{ij}$ denotes the
private cost to individuals engaged in transportation of an additional unit
shipped if use of roads and intersections is free. The cost actually in-
curred by all users of the road network exceeds this amount by $r_{ij} + r_i + r_j$
and is borne by traffic participants other than the one who caused this cost.
The analysis of the preceding section has shown that as the physical
capacity decreases relative to the amount of traffic programmed, that is
to say as congestion increases, the gap between private and social cost
$\frac{r_i + r_j + r_{ij}}{\frac{dc_{ij}}{du_{ij}}}$ is widened. The toll term $r_{ij}$ is capable of a direct inter-
pretation in the light of the definition (3.4). The ratio $\frac{dk_{ij}}{du_{ij}}$/
$\frac{dc_{ij}}{du_{ij}}$

$\frac{dc_{ij}}{du_{ij}}$ denotes the increase in cost which is caused to a unit of flow by a
unit increase in total flow. The term $p_{ij} = c_{ij} \frac{dk_{ij}}{du_{ij}}$ therefore
measures the total additional cost suffered by traffic on the road from a unit increase of flow.

In the case of fixed speed $r_{ij}$ was found to be the opportunity cost with respect to traffic excluded from this road (p. 16). We conclude that the efficiency conditions (3.5), (3.6) stipulate the equality of both costs, that to traffic on the road and that to traffic excluded from the road. This may be considered the marginal cost principle for the costs of congestion.

4. A Note on Computation.

4.1. Electrical Analogues. Computationally, the present problem may be characterized as a nonlinear saddle value problem, but the considerable number of variables involved renders rather infeasible the standard methods of calculating saddle points, as developed for the solution of 2-person games. In the case of one commodity the most elegant way of finding the optimum flow distribution appears to be by electrical analogue. This analogy is based on the observation stated in Kirchhoff's laws, that in an electrical network with given sources and sinks, the flows of current are such as to minimize total energy consumed.

If $R(I)$ denotes the electrical resistance of a wire as a function of amperage, then the energy required to transport a current of strength $I$ through this wire is $R(I)I^2$. Suppose now that the resistance of the wire from $i$ to $j$ is

$$R_{ij}(I) = \frac{k_{ij}(u_{ij}(I_{ij}))}{I_{ij}},$$

when $I_{ij}$ denotes the current from $i$ to $j$. Then the resulting currents will be such as to minimize

$$\sum_{ij} R_{ij} I_{ij}^2 = \sum_{ij} k_{ij}(u_{ij}(I_{ij})) I_{ij}$$
subject to the given sources and sinks. If then these sources and sinks are adjusted to equal the originations and terminations prescribed by the transportation program, it is seen that the electrical flows \( I_{ij} \) are equal to the optimal commodity flows \( f_{ij} \).

Since the output of an electrical source depends however on the potential, it is necessary to vary experimentally the voltage of the sources assembled for each input (output) until the resulting potential distribution gives rise to exactly the desired input (output). These potentials measure the efficiency prices at each location [1, Section 2.11, pp. 258–259]. Suppose now that in an analogy network which is to feature several commodity flows we superimpose the systems of electrical sources and sinks representing the various commodity originations and terminations. The resulting current will deviate from a pattern possible with commodity flows, since the electrical currents can take a number of short cuts through the different source and sink systems which have been contacted at various origination and destination points. In the absence of a means of eliminating these disturbances completely, the obvious device for keeping down cross currents of this kind is to install between the sources and the input contacts (sinks and output contacts) electrical resistances which are of an order large as compared to the network resistance. Then only a small proportion of current will take its way from one source system to another since that involves passing these strong resistances twice. The size of these resistances may be determined by experiment.

To ensure that currents in different directions will not cancel, a pattern of "one way traffic" in each wire representing a lane must be enforced by the use of rectifiers. As in the case of a one-commodity model the voltages of the sources and sinks that will produce the prescribed outputs and
inputs representing the originations and terminations of the respective commodities, must be found by experimentation. The net advantage of an electrical analogue over a trial and error method of computation therefore lies mainly in the easier control and quicker adjustment that may be had in the physical manipulation of the former.

In the case of fixed capacities the standard methods of linear activity analysis are applicable but, at the same time, are likely to be cumbersome since our problem involves so many variables. Instead of discussing these standard techniques, we shall outline here a method of successive approximation which leads to the solution in finitely many steps.

4.2. Computation. The idea of this method is to assign traffic first to routes which would be efficient in the absence of capacity limitations and to bring about the proper adjustment to capacities by successive impositions of efficiency tolls.

Efficient routing without regard to capacity can be considered a solved problem: it is a Koopmans-Hitchcock transportation problem, for which the simplex method has been well developed [8]; in the case where each commodity has one origin and one destination only, the cheapest routes may be found directly by looking at the highway map.

In the following what is said for roads applies equally to intersections. Generally a reduction of flow for a given road is to be carried out in two steps. 1) If there exists an equally cheap alternative route, i.e., if the road is part of a neutral circuit (n.c.) [p. 17] of which the route contains a whole branch, flows may be shifted among branches without change in cost. 2) Otherwise a toll can be imposed so as to create a neutral circuit and make the first operation possible. The manipulation of tolls is subject to two constraints: a) If flow on a road drops below capacity, the
toll must be lowered to the point where it equates capacity and flow, by making that road part of a neutral circuit as above (2), or otherwise to zero. -b) A change in toll on any road that is part of a previously established n.c. must be accompanied by the same change in toll on all branches of the n.c.

The natural order in which to consider the different roads is that given by the levels of the excess of flow over capacity. As soon as this excess has been reduced for one road below the level of another road, the operation proceeds on the other one. In case of a tie between several roads any one may be considered next. It is never necessary to consider several roads simultaneously because of the precaution in condition b). If this order of computation is adhered to it follows that the level of excesses of flow must decrease uniformly on all roads, and hence the process must converge in finitely many steps.

It should be noted that this computational method assumes that determination in each case of nearest detours (n.c.'s) is routine and requires no systematic attention.

The details are perhaps best demonstrated by a simple example.

4.3. An Example.

Figure 5.
In Figure 5 the circled numbers denote the locations, the other numbers transportation costs. Table I lists the traffic (aggregated for both directions) between each pair of origins and destinations; Table II contains the capacities (intersection capacities in the diagonal). Tables II include one table for each intersection of limited capacity (only that of point 2 in this example) and one for each road. The upper left field of each table shows the index of the road or intersection. In the

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>9</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table I
Transportation Program

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>15</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0</td>
<td>10</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Table II
Road and Intersection Capacities

Tables III:

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>-3</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>35</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>36</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>-2</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>45</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Excess</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roads</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>-1</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>-2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>-3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excess</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>13</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>4</td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excess</td>
<td>-6</td>
<td>-4</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>23</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>9</td>
<td>-2</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>3</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>2</td>
<td>-2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excess</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>24</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>6</td>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>-2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excess</td>
<td>-1</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>46</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>-----</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>46</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>36</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td></td>
<td>2</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Excess</td>
<td>-7</td>
<td>-5</td>
<td>-4</td>
<td>-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>----</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>empty</td>
</tr>
<tr>
<td>Table IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road or Intersection</td>
<td>12</td>
<td>12</td>
<td>23</td>
<td>12</td>
</tr>
<tr>
<td>Toll</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

The column below are listed the indices of pairs of origin and destination of flows which might pass this road. No distinction is made between directions; 34 stands also for 43.

These index pairs are entered as the flows happen to be assigned to this road. The next column bears the number 0, signifying the zero step of the analysis, which is the routing of traffic over the cheapest paths without regard to capacities. For instance flows 16 are routed over roads 12 and 26. An entry for this is made in the 0 columns of the tables for roads 12 and 26. For flows 36 we have a neutral circuit, composed of the routes 32, 26 and 34, 46. We arbitrarily assign half of this flow to either road. At intersections the originating and terminating flows are fixed by the program and may be disregarded in the tables.

If all flows have been assigned to roads the zero column is summed and the capacity figure subtracted from this. The result, entered as "excess" in the last row of the zero column, determines the order of the next step, which in this example is concerned with intersection 2. Apparently the only existing n.c. through 2 is 32, 26; 34, 46. The 2 units of flow 36 may
now be shifted from roads 32, 26 to roads 34, 46 relieving the excess flow through 2 by two units. The bookkeeping of this transfer is shown in columns 1 in the tables for intersection 2 and roads 23, 26, 34, and 46. The excess flow through intersection 2 being reduced to 5, 12 becomes the road of highest excess. The nearest n.e. involving this road, 12, 24; 13, 34 requires the imposition of a toll $r_{12} = 2$. This toll is entered in Table 4, and the transfer of 2 units of flow 14 from branch 12, 24 to branch 13, 34 is marked down in columns 2 of the tables for intersection 2 and roads 12, 24, 13, 34. With an excess of 4 units of flow, road 12 is still leading. The next neutral circuit 12, 26; 13, 34, 46 is set up by raising the toll on road 12 from 2 to 3. Shifting 3 units of flow to the branch 13, 34, 46 reduces the excess on road 12 to 1. With an excess of 2, road 23 is next. A toll $r_{23} = 3$ permits the rerouting of 2 units of flow 23 via roads 34, 42. This is step 4. There remains an excess of 1 on road 12. An increase of $r_{12}$ by 1 opens route 12, 23, 42 to flows 12 or 15. But to maintain the neutral circuit 23; 34, 42 the toll/23 must be raised simultaneously by 1. This final step 5 is entered in Table IV and shown in columns 5 of Tables III. The excesses in the last columns are now all non-positive; no further rerouting is required. Horizontal addition in all Tables 3 now yields the final flows on each road.

The graphs of all flows are assembled in the various parts of Figure 6.
Conclusion. The cost of the congestion generated by each traffic participant is a well defined magnitude composed of: the cost caused to other traffic on that road by a deterioration of speed and the cost of detours forced on traffic that would otherwise have used that road. An efficient combination of transportation activities requires 1) that with given origin and destination the cost of transportation, including that of the congestion created is equal on all routes used and not greater than costs on routes not used. 2) that the two components of the cost of congestion are equal at the margin. In the case of many origins and destinations for a commodity
(e.g., empty equipment), there is required in addition the existence of shadow prices on the commodity locations such that the cost of transportation; including that of congestion, is equal to the geographical price difference on all routes actually taken and greater or equal over any other part of the network. In the application of these criteria to each traffic element the capacity restrictions can be disregarded since they are fully expressed in the cost of congestion. To put the same fact differently, with each congested road and each congested intersection can be associated a shadow toll such that calculation of the optimal routes for each element of traffic can be carried out without regard to physical limitations, by including these tolls in the costs of transportation.

While all this is, economically speaking, common sense, the following result has been more difficult to establish and is valid only under special assumptions concerning the shape of the road capacity functions and the input functions for transportation resources (transportation cost functions), the assumptions made in this paper being plausible however from a technological point of view. This conclusion is that the conditions stated as necessary are also sufficient as a criterion of optimum allocation of the given network capacity. In particular this means that it is possible to find efficiency prices that contain all the information required for individual decision-makers to arrive at a socially efficient combination of transportation activities. To clinch this important point, a description will be given of an (intellectual rather than practicable) allocation game. This game is copied in all essential features from Koopmans' Analysis of Production [4, pp. 93, 94], except for the rule by which speed is determined on each road.
The participants in this allocation game will be called **traffic coordinators** (one in charge of traffic on each road and one in charge of each road intersection), **custodians** (in charge of roads, intersections and other resources), **shippers** (in charge of transportation activities), and **program managers** (one in charge at each terminal). The rules of conduct as prescribed by (1.1)...(1.5), (2.1)...(2.4) and (3.6) are the following.

For shippers: Trade and ship commodities between different locations whenever such activity yields a nonnegative profit. Profits are to be calculated as the geographical price difference minus cost of transportation resources including tolls (per commodity unit) for the use of roads and intersections. Do not engage in transportation activities with negative profits.

For the program managers: At origination points of a commodity sell the quantities specified by the transportation program at the highest prices for which shippers can be induced to take them over. At termination points of a commodity, buy the quantities required by the transportation program at the lowest prices for which they can be obtained from shippers.

For the custodians of resources other than roads and intersections: Charge market prices or shadow prices expressing the relative scarcities of these resources with regard to the demands of the rest of the economy. Be prepared to lower your prices to zero if available stocks exceed demand.

For the custodians of roads and intersections: Charge tolls just high enough to adjust the demand for passage to the road (intersection) capacity. Charge no tolls if the demand does not reach the capacity limit.

For the traffic coordinators: Make levies of a fixed proportion to profits on shippers and road custodians, paying back a similar proportion in the case of losses. Regulate speed so as to maximize expected additional income.
from these levies calculated on the following basis.

To calculate the expected increase in profits of shippers assume expected shipments to be equal to present ones and let expected cost be based on present tolls and prescribed speeds. To calculate the expected increase of toll incomes take present tolls and derive expected traffic flows by inserting prescribed speeds into the capacity functions.

In other words, if

\[ u_{ij} = \text{present speed} \]
\[ f_{ij} = \text{present flow} \]
\[ \hat{u}_{ij} = \text{prescribed speed} \]
\[ r_{ij} = \text{present toll} \]
\[ a = \text{proportion of profits taxed} \]

then the rule requires that

\[
\max_{\hat{u}_{ij}} a \left[ f_{ij} \left( k_{ij}(u_{ij}) - k_{ij}(\hat{u}_{ij}) \right) + r_{ij} \left( c_{ij}(\hat{u}_{ij}) - c_{ij}(u_{ij}) \right) \right]
\]

be found. This involves the capacity and cost function of only one road.

The conclusion of this paper is that under these rules an efficient combination of activities once established will be preserved. In other words, once the efficiency prices and speeds are computed traffic can be coordinated without direct control by communicating information about these tolls and speeds to all participants.

Instead of directly computing the speeds and prices, perhaps the game may be set up in such a way as to lead automatically to the attainment of the solution. But this raises serious problems of stability that cannot be treated here. If we wish, we may identify the players in this allocation game with entrepreneurs and the efficiency prices with market prices which would result under the reign of perfect competition. All roads would be
toll roads owned or at least operated by private firms (although some roads might yield zero tolls). But practical difficulties would arise, not only in the collection of tolls efficient enough to prevent new congestion, but also in the enforcement of competitive behavior by road custodians and by traffic coordinators as required by their role of selfless marginalists.

The present study may serve as a master pattern for the analysis of more general network problems involving, e.g., the time component of variable programs. Some extensions are straightforward, such as the admission of several types of equipment with different cost characteristics (leading perhaps to a separation of slow from fast traffic on various roads), or the replacement of one way roads by two way roads with the possibility of continuous substitution of flows (the tolls $p_{ij}$ and $p_{ji}$ must be equal).

The replacement of uniform speeds by speed distributions will be the subject of a separate paper.

Additional problems are raised if transportation equipment is also subject to availability restrictions. In the context of a dynamical network model some interesting results may be expected concerning the accumulation and decumulation of equipment at various locations.

2. The author has benefitted from comments by I. N. Herstein, H. Markowitz, and M. Slater. His main indebtedness is to T. C. Koopmans, who has suggested this problem, stimulated its treatment in various discussions, and presented an earlier version of it at the Logistic Conference, January 1952, Washington, D. C. The bibliographical and literary assistance of C. B. McGuire is gratefully acknowledged.

3. For convenience we shall speak of a highway network although some of the considerations seem to be more widely applicable.

4. Here the congestion caused by and the inefficiencies connected with the actual collection of tolls have been disregarded. To that extent the advantages of a toll system are, of course, overstated.


6. A function is called convex if the chord spanned by any two points on its graph lies above the graph. Analytically: \( f(x) \) is convex if
\[
\lambda f(x_1) + (1- \lambda) f(x_2) \leq f(\lambda x_1 + (1- \lambda) x_2)
\]
for all \( \lambda \) such that \( 0 \leq \lambda \leq 1 \). A function is concave if its negative is convex.

7. If some of these resources are subject to limitations these restrictions will, in general, conflict with the road capacity constraints in such a way as to render the marginal cost conditions insufficient for efficiency. The scope of the present paper does not permit our going further into this, but the study of Section 3 will convince the reader that the persistence of relative convexity of inputs with respect to flows is highly sensitive to the nature of the restrictions imposed.

8. The main obstacle to a direct application of known results in nonlinear activity analysis \([5] [6]\) is the failure of the minimand function \( \mu(\varphi, \lambda) \) to be convex jointly in the variables \( \varphi \) and \( \lambda \).

9. The vector \( \mathbf{v} \) is preassigned and not a function of the \( c_i \) (cf. p. 15).

10. If variable programs are admitted, there is a third component: cost due to the failure of traffic to arise that otherwise would have been generated.
References


