Transpn Res. Vol. 8. pp. 11-27. Pergamon Press 1974. Printed in Great Britain

TRANSPORT NETWORK OPTIMIZATION IN THE DUTCH INTEGRAL TRANSPORTATION STUDY

PETER A. STEENBRINK
Netherlands Railways, Utrecht, The Netherlands

(*Received* 4 *August* 1972, *in revised form* 10 *April* 1973)

Abstract—In the first part of this paper, the framework of the Dutch Integral Transportation Study is given, and the position of the defining of the road network is pointed out. In the second part, the transport network optimization problem is stated formally and the algorithm devised to solve this problem is described. The problem is formed by the minimization of the total costs (investments and users costs) for the transport of a given tripmatrix by choice of the dimensions and flows of the links of a transport network. The solution method has been devised especially in order to overcome the computational difficulties at large networks. It gives a reasonable solution in a relatively very short computation time. To that end the problem is decomposed into a number of sub-problems, yielding the optimal relationship between the dimension and the traffic flow for each link, and a master problem assigning the flows to the network in such a way that the objective function is minimized. In the third part the application of this method to the optimization of the Dutch road network is described.

INTRODUCTION

In 1967 the Dutch Minister of Transport charged the Netherlands Economic Institute with an integral transportation study for the Netherlands. The final report of this study has been published now (Nederlands Economisch Instituut, 1972). The aim of the study was to forecast the need for interurban transport infrastructure in the Netherlands for the period from 1970 to 2000. To that end the demand for transport has been determined and next the transport networks meeting this demand in the best way have been defined. This "best" is then expressed in terms of social objectives.

This paper focuses on the method used for the defining of the road network. In this the network is sought that minimizes the total social costs of transportation for a given matrix of car trips. The method described in this paper yields a reasonable solution to this problem in a relatively very short computation time. The paper consists of three parts. In the first part the position of the road network optimization in the whole transportation study is pointed out and it is argued why network optimization has been used. The second part gives a general description of the transport network optimization problem and the method devised to solve the problem. In the third part some practical experience at the Dutch Integral Transportation Study is reported.

PART I. THE DUTCH INTEGRAL TRANSPORTATION STUDY

Framework of the Dutch Integral Transportation Study

The spatial structure and the transport infrastructure are very closely related to each other and they form in fact one system. So the optimal solution to this one system has to be found. In this the maximization of the support to the well-being of the nation forms the objective. However already the quantification of this objective is so difficult that the problem had to be stated in a simpler way. The optimal transport infrastructure has been sought for a given land-use. Originally it was intended to solve this problem for different alternatives for the land-use. Cost and time constraints at the study forced to restrict to only one land-use alternative.

The objective in the defining of the transport infrastructure is again the maximization of the support to the well-being of the nation. To give concrete form to this objective the effects of transportation are divided into benefits and costs.

The fact that the transfer of persons and goods is possible and takes place is considered as the benefits, while the costs of constructing and maintaining the infrastructure together with the costs of using it, consisting of time consumption by the travellers, vehicle-operating cost, cost of accidents, damage to the environment and so on, are considered as costs. The difference between these benefits and costs, which are benefits and costs to the society, forms the objective function to be maximized. However again the working-out of this objective function raises many difficulties. This is especially the case with the evaluation of the benefits, which is closely related to the evaluation of the land-use. To avoid these difficulties the demand for transport has been defined in advance and has been supposed to be independent of the ultimate networks. One may consider this given demand for transport as the (constant) benefits of the transport infrastructure. The problem simplifies then to the minimization of the total social costs for a given tripmatrix.

The transport infrastructure problem is essentially a dynamic problem. The works constructed have a long life-time and must serve many situations changing over time and moreover other long-lasting effects exist. In order to take into account the dynamics of the system three comparatively static problems have been solved (the optimal networks for the transport of the years 1980, 1990 and 2000). Having these solutions, a coherent investment schedule for the whole period may be composed.

An optimal transport infrastructure implies a coherent system of road and railway networks. However, in connection with the charge of the study, the costs inside the towns have not been included into the objective function. Because these very costs play a decisive role at the choice between road and rail, it did not seem sensible to try to minimize directly the joint costs of road and rail transport.

Now an attempt has been made to get an estimate of the optimal coherent system of networks for the different transport modes by splitting the total transport up into the transport by mode in two ways: one with a high share of car traffic and a low share of rapid transit and the reverse situation. For both situations the minimal social cost networks are defined. Comparison of the results with respect to the computed social costs and with respect to estimates of factors not included into the objective function may show something of the overall optimal transport system.

Practically the above has been realized by constructing and applying a set of different models. The demand models for persons and goods transport as usually divided into production and attraction, distribution and modal split models form a very important part of this set (see the final report of the Dutch Transportation Study, Nederlands Economisch Instituut, 1972 and Hamerslag, 1972). These models (try to) describe the behaviour of the travellers and transporters in response on input situations. The input data for these models, among which the land use, the income and the car-ownership are most important, have been estimated for the 3 yr 1980, 1990 and 2000. Next the demand models have been applied, which resulted in 3×2 tripmatrices both for car traffic and rail transport.

Given the possibilities for the networks and the different cost elements depending on the dimensions of the links and the flows on them, the networks have been defined meeting these tripmatrices in the best way. To get the optimal railway network the method of Barbier (1966) extended by Haubrich (1972) has been used. For the road network the method described in this paper (see also Steenbrink, 1971 and 1974) has been devised and applied.

To give an idea of the size of the study: the study area, consisting of the whole Netherlands, has been divided into 351 transport-zones; while for the modal split and the railway-optimization it is worked even with about 1250 transport-zones.

The road network is given in Fig. 1. It consists of about 6000 links and about 2000 nodes. The railway network has a simpler structure, but because of the large number of stations considered (about 900) the number of nodes and links is still rather large and of the same order as for the road network.

Why network optimization?

As stated in the Introduction, the aim of the study is to advise on decisions to be made about the transport infrastructure. The decisions are made on account of the costs (and benefits) involved with them, so it is necessary to have available a set of models giving these costs (and benefits) in dependence on the decisions made. Many aspects of the costs are essentially dependent on the geographical location of the different roads with their dimensions and the traffic flows on them. So a model which computes the traffic flows in dependence on the network forms a necessary part of the transport model set. This "traffic assignment" model must meet the condition, that the results of the assignment process depend on the decisions made. So only the capacity-restraint techniques can be applied (including the diversion-type capacity-restraint techniques, at least when the necessary computation-time is reasonable).

A very common procedure now is to define a limited number of alternatives for the network and to compute the traffic flows and next the costs (and benefits) on these networks. On account of these results a certain network is chosen as the best one or some more alternatives are proposed and evaluated. For every flow computation it is necessary to put a fully dimensioned network into the computer. Though this input-network is usually defined very carefully and consciously on some places, it is defined very roughly and rapidly on other places. But still the results are influenced by the dimensions of all links of the network, and the fact that it is (almost) impossible to trace out these influences at the interpretation and evaluation of the results raises an important danger. Very common and trivial examples of this phenomenon are the cases in which a road, which is expected to be heavily-loaded, gets in advance a large dimension and so also a heavy traffic flow and conversely. This danger is the greater, because the arrangement of the input is not a very interesting job and so it tends to get too little attention.

Fig. 1. The road network used at the Dutch Integral Transportation Study.

The only way to meet these objections is to evaluate very many, or preferably all, alternatives for the network. But as we know, the number of alternatives is enormous. If only for 30 links two alternative dimensions must be considered, the total number of combinations is 2^{30} or about a thousand million. Very powerful techniques must therefore be used to overcome these difficulties in evaluating all alternatives and finding the best one. Therefore one will sometimes be content with a rather good, instead of the best solution. The algorithm described in this paper is developed with the very objective of minimizing the necessary computation time in the search for a good solution. The algorithm presented here provides a rather good network *in a computation time equal to the computation time necessary for the evaluation of one alternative.* It seems hardly possible to construct a faster algorithm.

The sensitivity of the results for the input exists also, of course, for the other parts of the transportation model, like the transport production and attraction, distribution and modal split. So far these parts too, it is necessary to evaluate all alternatives. As said in the preceding section, until now this seems to have been a problem too difficult to solve, at least for large-sized transportation-systems. Fortunately the results of the other transport models seem to be less dependent on the input, sometimes, and on the other hand, the number of possible alternatives is not always very large. So, although it is necessary to solve the overall transportation optimization problem, it seems to be reasonable to start with the optimization of the assignment/infrastructure-network part.

PART II. THEORETICAL TREATMENT OF THE PROBLEM AND THE SOLUTION METHOD

The transport network optimization problem: statement of the problem

The objective function to be minimized consists of all costs of transport for the society. These costs consist of cost related to the construction and maintenance of the transport infrastructure and of users cost. We suppose that these costs are defined on the links of a transport network.* We suppose further that the costs for a link are a function of the traffic flow, the dimension and the characteristics of the link:

$$F_{ij} = F_{ij}(c_{ij}, x_{ij})$$

with F_{ij} = total cost on link ij
c_{ij} = dimension of link ij (capacity, number of lanes)
x_{ij} = traffic flow on link ij.

The objective function now is formed by the summation of the costs over all links:

$$F = \sum_{ij \in L} F_{ij} \tag{1}$$

with L set of links.

The network is the most important *decision-variable*, the structure and the dimensions of the network must be defined. However, the structure of the network is considered as given and it is concerned with the dimensions of the links only. Because the dimension zero is included as possibility, in fact a structure is chosen among a set of possible structures.

* As usual, a network is defined as a set N of elements, called nodes, and a set L of ordered pairs of these nodes, called links. In a transport network the links generally represent roads and the nodes intersections of these roads, centroids of traffic zones (origins and destinations) or other points of interest. A flow is defined by the relationship in equation (2) and represents the number of persons or vehicles travelling on a road.

For the determination of the traffic flows there are two possibilities (see, for instance, Wardrop, 1952, and Beckmann, McGuire and Winsten, 1956). The first one is the descriptive assignment, in which the traffic flows are defined according to a model describing the individual route choice as well as possible. In that case the route choice model forms part of the constraints. The second possibility is the normative assignment, in which the traffic flows are chosen in such a way that the objective function is minimized. We will use here the normative assignment, which makes it easily possible to use the decomposition described in the next section. Other possible decision-variables, such as the traffic management system or the road-pricing system, will not be considered here.

The most important *constraints* are formed by the given tripmatrix and the fact that the traffic flows are conservative in every node except for the origins and destinations, and are never negative and additative to each other on a link:

$$\sum_{\substack{i \\ (ij \in L)}} x_{ij}^{ab} - \sum_{\substack{k \\ (jk \in L)}} x_{jk}^{ab} = \begin{cases} 0; & \text{for all } j \neq a \text{ or } b;\ j \in N \\ -x^{ab} & \text{if } j = a \\ x^{ab} & \text{if } j = b \end{cases}$$

for all $ab \in P$

$$x_{ij}^{ab} \geqslant 0 \text{ for all } ij \in L; \quad \text{for all } ab \in P$$
$$x_{ij} = \sum_{ab \in P} x_{ij}^{ab} \text{ for all } ij \in L \tag{2}$$

with x^{ab} = number of trips from a to b
x_{ij}^{ab} = number of trips from a to b flowing on link ij
N = set of nodes
P = set of transport relations (origin–destination–pairs).

Moreover the dimensions for each link are constrained:

$$c_{ij}^{\min} \leqslant c_{ij} \leqslant c_{ij}^{\max}; \quad \text{for all } ij \in L. \tag{3}$$

The lower bound will be mostly the existing dimension of zero, while the upper bound can be set because extension beyond this bound is physically, technically or environmentally impossible or too expensive to be interesting.

So the problem is formulated as follows:

$$\min_{c_{ij}, x_{ij}^{ab}} \sum_{ij \in L} F_{ij}(c_{ij}, x_{ij}) \tag{4}$$

subject to equations (2) and (3).

Or in vector notation:

$$\min_{c, x} F(C, X) \text{ subject to } X \geqslant 0,\ AX = 0,$$
$$C^{\min} \leqslant C \leqslant C^{\max}.$$

Some remarks on the well-known solution methods

The cost minimization problem for the normative and the descriptive system as well has been attacked by many researchers. General surveys of possible solution methods are given by Manheim (1968), Bergendahl (1969) and Steenbrink (1974). The constraints of equations (2) and (3) are all linear. Expressing the objective function as a piece-wise linear or a quadratic function linear respectively quadratic programming can be applied for the normative system. However for realistic networks the number of decision variables and constraints is so large, that it seems impossible to use these methods in practice, even using techniques to reduce the size of such programs, as for instance proposed by Tomlin (1969). Because of the combinatorial nature of the problem, the use of combinatorial techniques as branch and bound has been proposed and applied (Ridley, 1965; Ochoa-Rosso and Silva, 1968). Mostly the minimization of the investment cost or the users cost with a descriptive assignment is treated, though it is possible to extend the statement of the problem. The size of the problem is decisive, however, for the usefulness and it does not seem possible to use branch and bound for really large networks. The use of heuristics to get a rather good solution instead of an optimal one in a computation time, which is also reasonable for large networks, seems more promising. However it can be very dangerous, too, and one must be very conscious of this at the interpretation of the results. The method of Barbier (1966), for instance, used for the defining of the railway networks at the Dutch Integral Transportation Study is an example of a heuristic technique. Finally, the use of aggregation and decomposition techniques seems very useful in efforts to overcome the difficulties caused by the size of the problem.

Concluding, at the moment of the design of the search and evaluation process at the Dutch Integral Transportation Study (begun 1970) it did not seem possible to make straightforward use of existing optimization techniques and to my knowledge this is still impossible today. This formed the reason for us to devise and apply the method described in the following sections.

Decomposition of the original problem

The basic principle of the method described in this paper is the decomposition, in which, as it were, one set of decision-variables (C) is eliminated. The original problem (4) is decomposed into n_L *subproblems* (with n_L being the number of links):

$$\left.\begin{aligned} &F_{ij}^{\min}(x_{ij}) = \min_{c_{ij}} F_{ij}(c_{ij}, x_{ij}) \\ &\text{subject to } c_{ij}^{\min} \leqslant c_{ij} \leqslant c_{ij}^{\max} \end{aligned}\right\} \text{for all } ij \in L \qquad (5)$$

and the *master-problem*:

$$\min_{x_{ij}} \sum_{ij \in L} F_{ij}^{\min}(x_{ij}) \qquad (6)$$

subject to $A\,X = 0$, $X \geqslant 0$.

The subproblems yield the optimal relationship between traffic flow and dimension for each link. These problems are mathematically very easy to solve. The inspection of all possibilities, if desired in combination with a search method, or differentiation if we assume $F_{ij}(c_{ij})$ to be a differentiable function of the continuous variable c_{ij} will yield the correct solution. The master-problem is mathematically much more complex. We will treat that in the next two sections.

First we prove that the combination of the master-problem and the sub-problems has the same solution as the original problem. Let us suppose F^* being the optimal solution to the master-problem. Then the following relationship is satisfied:

$$F^* \leqslant \sum_{ij \in L} F_{ij}^{\min}(x_{ij}) \text{ for all } X$$

From the statement of the sub-problems however it follows that for every particular solution X' holds:

$$\sum_{ij \in L} F_{ij}^{\min}(x'_{ij}) \leqslant \sum_{ij \in L} F_{ij}(c_{ij}, x'_{ij})$$

for all c_{ij} subject to $c_{ij}^{\min} \leqslant c_{ij} \leqslant c_{ij}^{\max}$.

Combining these two relationships it is seen that F^* is the minimum solution to the original problem and the decomposition is correct.

Conditions for the optimal solution of the master problem

Before going into the solution methods for the master-problem, we will first derive the necessary and sufficient conditions for the optimal solution. This is easily possible for the case of convex cost functions. So first we will discuss some of the properties of the objective function per link. The users cost (time consumption, vehicle operating cost, accidental cost) will be convex in general due to effects of congestion. The investment cost is in the first place mostly discrete. Moreover for new roads the initial investments will generally be rather high, the investments for further expansion will be generally lower until a certain expansion point and become high again beyond this point. So generally the objective function will be of the form of Fig. 2 with "switching points" from n to $n + 1$ lanes. This function is certainly non-convex. But it is possible to use an "envelope curve" as in Fig. 3 or to assume the dimension a continuous variable. Both operations do remove much of the non-convex nature of the objective function. In fact the functions become convex above a certain value for the traffic flow, which area is the most

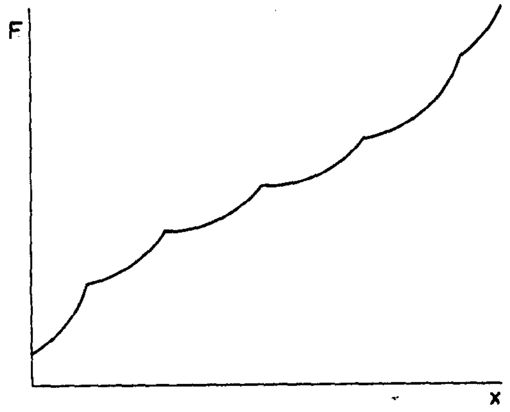


Fig. 2. Objective function on a link

relevant area indeed. Moreover when the road has an existing dimension sufficiently large the cost function is convex on the whole area. Research in the investment cost for the Dutch highways showed that using continuous dimensions for the roads, the objective function is convex for roads with four or more lanes and that the concavity for the first interval is not very serious.

It seems reasonable to derive conditions for the optimal solution for the case of a convex objective function. Similar derivations have been given by Beckmann (1956), based on the optimality conditions of Kuhn and Tucker (1951)—Murchland (1969) drew attention to this work—and by Dafermos and Sparrow (1969). We will use here an approach suggested by Timman (1966).

The problem is stated as follows, using the notion of paths:

$$\min_{x_{ij}^{pab}} F(X) = \sum_{ij\in L} F_{ij}^{\min}\left(\sum_{\substack{p\in Pa^{ab}\\ ab\in P}} x_{ij}^{pab}\right) \tag{7}$$

subject to:

$$x_{ij}^{pab} = x_{kl}^{pab}; \quad \text{for all } ij\in p \text{ and } kl\in p, \ \text{for all } p\in Pa^{ab}, \ \text{for all } ab\in P \tag{8}$$

$$\sum_{p\in Pa^{ab}} x_{aj}^{pab} = x^{ab}; \quad \text{for all } ab\in P \tag{9}$$

$$x_{ij}^{pab} \geqslant 0; \quad \text{for all } ij\in L, \ \text{for all } ab\in P, \ \text{for all } p\in Pa^{ab} \tag{10}$$

in which: x_{ij}^{pab} = number of trips from a to b, using path p from a to b, flowing on link ij

Pa^{ab} = set of paths from a to b.

Suppose X^* is the optimal solution, then the following relationship holds:

$$F(X^*) \leqslant F(X^* + \Delta X) \tag{11}$$

in which $X^* + \Delta X$ is a feasible solution in the "neighbourhood" of X^*.

Because $X^* + \Delta X$ is feasible the following relationships hold for Δx_{ij}^{pab}:

from (8)

$$\Delta x_{ij}^{pab} = \Delta x_{kl}^{pab}; \quad \text{for all } ij\in p \text{ and } kl\in p, \ \text{for all } p\in Pa^{ab}, \ \text{for all } ab\in P \tag{12}$$

from (9)

$$\sum_{p\in Pa^{ab}} \Delta x_{aj}^{pab} = 0; \quad \text{for all } ab\in P \tag{13}$$

from (10)

$$\Delta x_{ij}^{pab} \geqslant 0 \text{ if } x_{ij}^{*pab} = 0; \quad \text{for all } ij\in L, \ \text{for all } p\in Pa^{ab}, \ \text{for all } ab\in P. \tag{14}$$

Expanding $F(X^* + \Delta X)$ into a Taylor series we get:

$$F(X^* + \Delta X) = F(X^*) + \sum_{\substack{p\in Pa^{ab}\\ ab\in P\\ ij\in L}} \left(\frac{\partial F}{\partial x_{ij}^{pab}}\right)_{X=X^*} \Delta x_{ij}^{pab} \tag{15}$$

+ higher order terms.

Because F is a convex function of X, the sum of higher order terms is always positive. Let us suppose further that the values of the Δx_{ij}^{pab} are so small that the sum of the higher order terms is negligible (but still positive).

That means that we can write the inequality (11) as:

$$\sum_{\substack{p\in Pa^{ab}\\ ab\in P\\ ij\in L}} \left(\frac{\partial F}{\partial x_{ij}^{pab}}\right)_{X=X^*} \Delta x_{ij}^{pab} \geqslant 0. \tag{16}$$

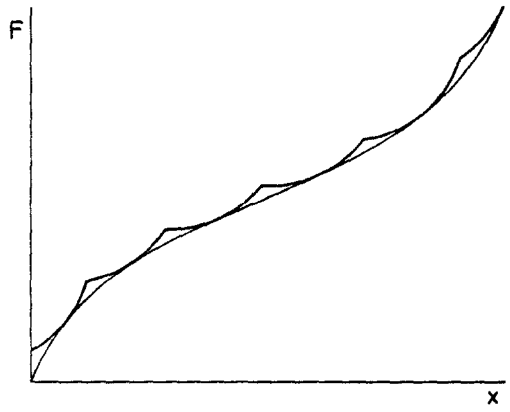


Fig. 3. Objective function on a link with an envelope curve

Substitution of the relationships (12), (13) and (14) into (16) and some working out gives:

$$\left.\begin{array}{l}\text{if } x_{ij}^{ab} > 0 \text{ then } F'^{ai} + F'_{ij} = F'^{aj} \text{ for all } ab \in P \\ \text{if } F'^{ai} + F'_{ij} > F'^{aj} \text{ then } x_{ij}^{ab} = 0 \text{ for all } ij \in L\end{array}\right\} \quad (17)$$

in which:

$$F'_{ij} = \left(\frac{dF_{ij}^{\min}}{dx_{ij}}\right)_{x_{ij}=x_{ij}^*} = \left(\frac{\partial F}{\partial x_{ij}^{pab}}\right)_{X=X^*}$$

$$F'^{ai} = \min_{p} \sum_{kl \in p} F'_{kl}; \quad p \text{ forms a path from } a \text{ to } i.$$

From the convexity of $F(X)$ we know that the necessary conditions of (17) are sufficient, and that the minimum is a global one.

Solution methods for the master-problem

The condition (17) for the optimal solution for a convex objective function resembles very much the second principle of Wardrop (1952) used to describe the individual route choice behaviour ("the journey times on all the routes actually used are equal and less than those which would be experienced by a single vehicle on any unused route"). The only difference is the use of the marginal objective function instead of travel time. But $(dF^{\min})/dx$ is a nonnegative increasing function of the traffic flow as t is. So it is obvious that we can use all well-known sophisticated traffic assignment procedures (see, for instance, Bruynooghe, Gibert and Sakarovitch, 1968, and Dafermos and Sparrow, 1969) to get the solution for the network optimization problem.

In the Dutch Integral Transportation Study the stepwise assignment is used, which is mathematically identical with the well-known capacity-restraint technique (Steel, 1965). The tripmatrix is assigned to the network part by part, in every step a part of the trips of each relation being assigned to the route with the minimum value for the marginal objective function on that moment:

$$\begin{array}{ll}\text{if } dx^{\alpha n}{}_{ij}^{ab} > 0 & \\ \text{then} & (F'^{ai})_{X=X^{\alpha n-1}} + (F'_{ij})_{X=X^{\alpha n-1}} = (F'^{aj})_{X=X^{\alpha n-1}} \\ \text{if} & (F'^{ai})_{X=X^{\alpha n-1}} + (F'_{ij})_{X=X^{\alpha n-1}} > (F'^{aj})_{X=X^{\alpha n-1}} \\ \text{then} & dx^{\alpha n}{}_{ij}^{ab} = 0 \end{array} \quad (18)$$

with: $X^{\alpha n-1}$ = traffic flows after the $(n-1)$th step
$dx^{\alpha n}{}_{ij}^{ab}$ = flow from a to b in the nth step assigned to link ij.

Although this process does not guarantee that the correct solution is found*—even when an infinite number of infinitely small steps is used—a reasonable solution is obtained in a simple way. Moreover the constraints are always satisfied.

It can be expected, and experiments show this, that the use of many steps give better results than the use of few steps. But the necessary computation time is strongly related to the number of steps, so one will try to use as few steps as possible. It is advantageous to use difference quotients instead of differential quotients. The criterion then becomes (ignoring further the division by the constant Δx):

$$F_{ij}^{\min}(x_{ij}^{\alpha n-1} + \Delta x) - F_{ij}^{\min}(x_{ij}^{\alpha n-1}). \quad (19)$$

The advantage of this is that we "look a little ahead", depending on the size of Δx. So for the assignment process the parameters n, α_n and Δx^n must be chosen. These parameters must be chosen in relation to each other in such a way that a reasonable solution is found in a reasonable computation time.

As said before, the conditions (17) are only sufficient or necessary for an optimal solution in the case of a convex objective function. Still, the stepwise assignment according to the least marginal objective function may also give correct solutions in the case of non-convex functions, dependent on the different data for the problem. For instance, if a certain road is always "cheaper" than another, the correct solution will be obtained. Moreover, sensible choice of the optimization parameters can be very profitable in such cases.

PART III PRACTICAL EXPERIENCE IN THE OPTIMIZATION OF THE DUTCH ROAD NETWORK

Concrete statement of the problem

For a concrete statement of the problem it is necessary to define exactly the objective function, the decision variables and the constraints. The objective function consisted of the following elements:

(a) evaluated travel time costs
(b) vehicle operating costs } users' costs
(c) costs of accidents
(d) investments and maintenance of the road.

All costs are defined on the inter-urban networks and for the traffic flows during working-days. The costs are all expressed in the same monetary units. The users' costs and maintenance costs are considered to be equal for every year (the traffic is considered to be constant,

* Only for the case of one relation and a network in which the different paths used do not contain the same links can it be proved that an infinite number of infinitely small steps gives the correct solution. In all other cases the following error can be made. The use of a certain link may be profitable for many relations at the beginning of the process. Due to the many trips assigned to that link and the high value for the marginal objective function caused by that fact, it may be possible that no relations at all use the link at the end of the process. A better solution might be obtained if some relations should have used the related link better, while other relations should have always used other routes (see Steenbrink. 1974, Sections 5.4.2 and 5.7.2).

namely the traffic for the years 1980, 1990 or 2000), while the investments are considered to be made in the related year itself. All costs are discounted to present value.

The users' costs are considered to be a function of the dimension and the traffic flow on the link. The costs of maintenance are considered to be dependent only on the dimension of the link, while the costs of investments are considered to depend on the (new) dimension and the existing dimension of the links and the nodes as well. The objective function is the summation of these costs over all links and nodes. Because the users' costs depend on the traffic flow and the traffic flow varies over the hours of the day and the days of the year, all costs vary over time. A certain distribution over time of the traffic has been supposed and the users' costs are expressed as a function of the flow on the average evening peak hour during working days.

The defining of travel time costs consists of defining of travel time as a function of the dimension and flow and of the social evaluation of travel time. The travel time function has been based on a relationship between mean speed and flow/capacity ratio comparable with similar relationships given in the Highway Capacity Manual 1965 (Highway Research Board, 1966) but based on Dutch data. The evaluation of travel time depends on the value of the times spent at other activities and the (dis)-pleasures of travelling. Because of this last fact the value of travel time has been considered also as a function of the traffic flow.

The social vehicle operating costs consists of fuel consumption, and so on (taxes not included) and that part of the depreciation that is directly caused by the use of the car. Also these costs have been determined as a function of the flow/capacity ratio. The material and evaluated non-material costs of accidents have been included into the objective function. Again these costs have been supposed to depend on the flow/capacity ratio.

All users' costs have been expressed in the mathematical form $x[\alpha + \beta(x/c)^{\gamma}]$ in which γ is 5. At this the values for α for the travel time costs and vehicle operating costs are almost equal and about 10 times as high as for the costs of accidents, whereas the value for β is the highest for the travel time costs.

For a number of different types of roads, intersections and special structures as bridges and tunnels the relationship between the construction costs and the number of (meeting) lanes has been defined. In these relationships the starting investments (until 2×2 lanes for a road) are generally rather high; expansion to 2×3 and 2×4 lanes is supposed rather cheap, while with further expansion to 2×5 and 2×6 lanes, again high investments are involved. The construction cost functions used are continuous functions of the continuous dimension and consist of three linear parts with different slopes as the first part goes through the origin. No construction costs are counted for the number of lanes already present. For all roads, intersections and other structures in the network it has been defined to which type they (will) belong. The costs for maintenance finally have been supposed also to depend linearly on the number of lanes.

The decision variables for the network optimization problems are the dimensions and the flows for all links. So a normative assignment has been used. (However see the next section for the final assignment used). The constraints were formed by the usual network constraints (2) and the dimension constraints (3). The tripmatrix included into the network constraints was one of the six forecasts of the total amount of traffic expressed in passenger car units between all 351 centroids of transport zones. The lower bound of the dimension equaled the existing dimension of a road in 1975 or zero in the case that no road was present in that year. For every road an upper bound had also been set. That meant that extension beyond this bound was technically or physically impossible or too expensive to be worth considering. Furthermore, the environmental aspects have been taken into account by constraints in such a way that some upper bounds are just set because further expansion would imply a serious damage to the environment. A general upper bound of 2×6 lanes has been used, but for many roads a lower upper bound has been applied. A large set of existing but not further expandable two lane roads has been specially introduced.

The resulting sub-problems had an optimal solution between 1100 and 1400 p.c.u. per lane during the average evening peak hour (for the normal case, without the influence of dimension constraints). This value fits very well with the existing ideas about the desired or maximally allowed, number of cars per lane during the peak hour. In fact, a small correction of the result originally found has been applied to improve this fitness.

With respect to the necessary (or desired) convexity of the objective function for the master problem it must be noted that all functions are convex for dimensions from 2×2 lanes. That means that the cost functions for all roads with 2×2 or more existing lanes are totally convex. For roads with no existing dimension the cost function is not convex for the whole dimension interval, but it is for the more relevant interval above 2×2 lanes. The convexity is especially clear on those places where dimension constraints cause that the flow is much above 1100–1400 p.c.u. per lane for the average evening peak hour.

The total computation process used

The optimization results in a dimensioned network and a traffic flow pattern, which is the result of a normative assignment. The real route choice behaviour will generally cause another flow pattern. So the use of a normative assignment is certainly an objectión. Three comments may be made about this:

(a) The normative assignment is socially optimal. One may wish to try to influence the individual route choice in such a way that the normative assignment and the descriptive one coincide as much as possible.*

(b) The route choice behaviour is not very well defined and may be influenced easily. So the assignment used does not matter too much.

(c) It is possible to start the computation process with a normative assignment in order to get the network in an easy way and to apply a descriptive assignment to the network found afterwards.

This last process has been used at the Dutch Integral Transportation Study. The total computation process has been started with the application of a stepwise assignment of the tripmatrix to the network according to the routes with the least value for the marginal objective function. This resulted in a dimensioned network. In the second phase of the process the tripmatrix was assigned to the network found with the stepwise assignment according to the least marginal objective function according to a descriptive route choice model.

In the third and last phase the dimensions of the links were adjusted to the traffic flows found in the second phase in such a way that the objective function was minimized.† A flow chart of the computation process is shown in Fig. 4.

There was a second reason for using the computation process mentioned. Because of the computation cost, attempts were made to use as few steps as possible

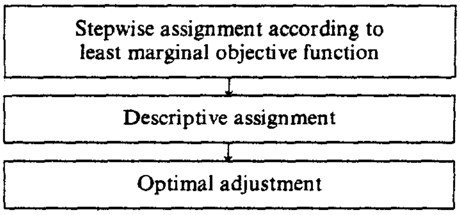


Fig. 4. Flow chart of the computation process used for the optimization of the Dutch road-network.

* This may be tried by traffic management or road-pricing. In the last case the charges must equal the difference between the marginal objective function and the total individual users costs to achieve an optimal solution.

† Iteration over the last two steps did not seem sensible because of the high computation cost involved.

at the stepwise assignment. Due to this and the connected choice of the parameters some "errors" arose. The most important one was that two lane roads (mostly already existing), which were not allowed to be expanded further, and roads inside towns, were not used well enough. These "errors" were reduced by letting follow the stepwise assignment according to least marginal objective function by a descriptive assignment and an optimal adjustment.

Parameters used

For the stepwise assignment according to least marginal objective function difference quotients have been used. Then the following parameters values had to be chosen:

the number of steps;

the size of the parts to be assigned to the network at the different steps (α_n, this can be different for every step);

the size of the difference at the marginal objective function (Δx^n, can be different by step too).

Moreover a number of links have not been included into the optimization process. The roads inside towns, for which no costs had been defined, and the centroid connectors belong to this category. For these links a value for the marginal objective function had to be put in. The values for the parameters had to be chosen in such a way that they provided a network in relation with each other as well as possible within a computation time as short as possible. An important factor was that one step took already about 12 min computation time on an IBM 360/65.

Experiments on a small network with a convex quadratic objective function showed it was possible to get a very reasonable solution with the use of a fairly small number of steps (generally four), provided that a proper choice had been made for the other two sets of parameters. Generally the best solution was obtained by applying equal steps. In that case the smallest largest step is obtained and the larger the step the larger "error" can be made.

Further a large value for Δx gave generally better results than a small value for Δx, caused by the fact that "errors" made at a large flow assigned are more serious than the ones made at a small flow assigned, and the larger Δx, the more sensitive the process is for large flows assigned. In the study for the Dutch road network not all functions were convex. Moreover the stepwise assignment according to least marginal objective function was followed by a descriptive assignment and an optimal adjustment, at which many "errors" could be "corrected". All these things had to be taken into account at the defining of the assignment parameters.

To be able to make a better choice, further experimentation has been made with different values for the parameters at the real network, the real objective function and a tripmatrix of the same order as the final ones and very similar to them. It turned out that the use of three steps was not advisable, because of the serious congestions appearing. Four steps seemed to be better and it was decided to apply four steps. These steps were chosen to be equal because of the arguments mentioned above. In the defining of the difference Δx it was considered that a heavily-loaded highway would have a traffic flow of about 6000 p.c.u. per direction for the average evening peak-hour. Because of the preference for large values for Δx, the value 2000 was chosen. That implied that some roads should not be loaded sufficiently. This was especially true for non-expandable two lane roads. However as mentioned before this unfavourable effect was reduced considerably by the following descriptive assignment and optimal adjustment. For the links not included in the optimization process, values for the marginal objective function have been put in. These values were chosen so that these links disturbed the process as little as possible. That implied for the roads inside towns per kilometer a value equal to the marginal objective function of an overloaded newly to construct circular road per kilometer, that is a rather high value.

For the centroid connectors a high value for the link as a whole has been taken.

For the following descriptive assignment a capacity-restraint process has been used. In this four steps have been used, assigning subsequently 0·1, 0·3, 0·3 and 0·3 parts of the tripmatrix to the network. The first step is so small because some roads may have got a very small dimension in the preceding stepwise assignment according to least marginal objective function and so can bear only a small flow, while they have still a low value for the travel resistance in the first step. The deterrence function has been chosen further so that the overall result gives a good simulation of the flow pattern resulting from the real route choice behaviour.

Finally, special values also had to be inserted here for links not included in the optimization process.

Illustration of the operation of the method at a part of the network

In order to give a better idea of the operation of the method we will discuss the intermediate results (after the different steps) of the application of the method to the road network of the Netherlands* for a small part of this network. We will show the traffic flows on the average evening peak hour and the dimensions of the six possible roads north of the Hague at the end of the four steps of the stepwise assignment according to the least marginal objective function (SALMOF) and at the end of the last step of the following descriptive assignment and optimal adjustment (see Table 2 and Fig. 5). For reasons of simplicity we number the roads 1–6 inclusive. Roads 1 up to and including 4, serve especially the town the Hague, while roads 5 and 6 are also important for through-going traffic. The dimension-constraints of the six roads are shown in Table 1. The investment-cost-characteristics for the roads 1 up to and including 4 are identical, while the roads 5 and 6 are a very little cheaper in constructing and/or expanding.

Table 1. Dimension constraints for the six roads

Road no.	In 1975 existing number of lanes	Maximal number of lanes
1	—	2 × 6
2	2 × 2	2 × 3
3	2 × 2	2 × 3
4	1 × 2	1 × 2
5	2 × 2	2 × 6
6	—	2 × 6

For all roads it holds that the marginal cost for investing in the area from zero to four lanes is almost equal to that in the area from eight to 12 lanes and about three times the marginal investing cost in the area from four to eight lanes.

In the first step of SALMOF the new roads, for which investments are necessary, and the non-expandable two-lane-road get (almost) no traffic flows. Among the roads 2 and 3, road 2 gets almost all traffic flow. Because of the equal characteristics this fact must be explained by the demand for transport, which will be larger in the area of road 2. This is true indeed, for in an all-or-nothing assignment according to the shortest distance even road 1 gets the most traffic.

In the second step it would be necessary to expand roads 2 and 5. In part the investment cost for this does not outweigh the extra users cost (time consumption, vehicle operating cost and so on) caused by the use of a longer route. So road 3 attracts a lot of traffic flow. The new roads and the non-expandable one are still not interesting.

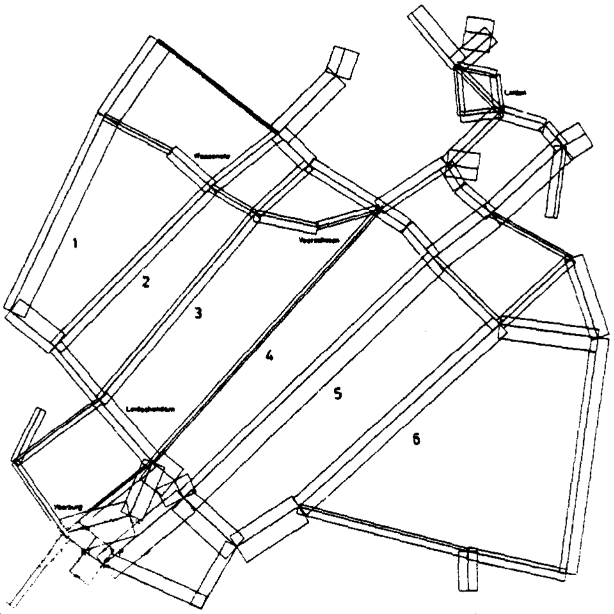
At the beginning of the third step roads 2 and 3 have been loaded heavily, especially from the city. Due to the dimension-constraint on these two roads, road 1 also gets a traffic flow. The overloaded road 3 gets no traffic at all from the city and road 2 gets very little. The existing roads to the city are expanded to three

* The results discussed here are taken from a test run and are different from any ultimate results of the Dutch Integral Transportation Study.

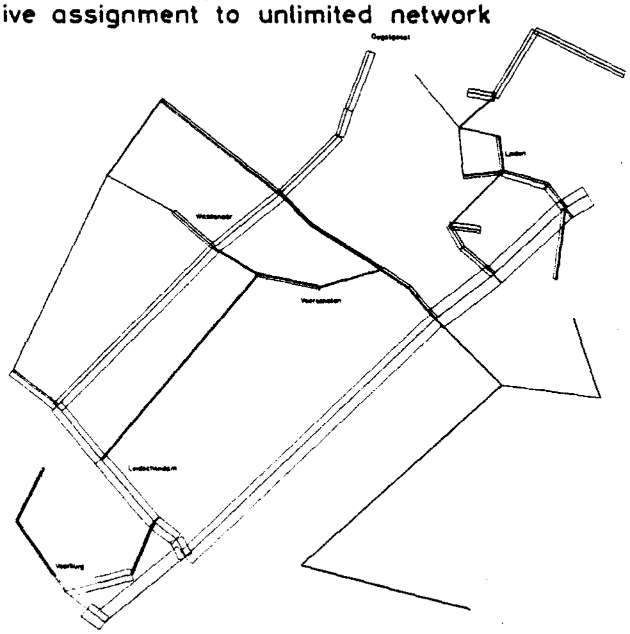
Table 2. Traffic flows (in evening peak hour) in p.c.u. and numbers of lanes for illustration of the method, six roads north of the Hague

Road	SALMOF 1st step		SALMOF 2nd step		SALMOF 3rd step		SALMOF 4th step		Ultimate solution	
	Flow	Dimension	Flow	Dimension	Flow	Dimension	Flow	Dimension	Flow	Dimension
1 to	25	0·0	117	0·1	386	0·4	917	1·0	2651	2·4
from	82	0·1	185	0·2	1438	1·3	3297	3·1	3874	3·6
2 to	1191	2·0	1686	2·0	2082	2·0	2352	2·1	2219	2·1
from	2184	2·0	2636	2·4	3218	2·9	3218	2·9	2955	2·5
3 to	167	2·0	2102	2·0	2827	2·6	3054	2·8	1161	2·0
from	333	2·0	3654	3·0	3654	3·0	3654	3·0	2713	2·5
4 to	0	1·0	0	1·0	0	1·0	0	1·0	820	1·0
from	0	1·0	0	1·0	67	1·0	196	1·0	1054	1·0
5 to	2171	2·0	2476	2·3	3712	3·4	4460	3·9	2767	2·5
from	2953	2·8	3772	3·5	5909	4·9	6716	5·5	5507	4·6
6 to	133	0·1	229	0·2	324	0·3	1512	1·4	1647	1·5
from	117	0·1	249	0·3	555	0·6	2367	2·2	2409	2·2
Total										
to	3687		6610		9331		12,296		11,265	
from	5669		10,496		14,841		19,448		18,512	

The road network north of The Hague



Descriptive assignment to unlimited network



SALMOF first step

Fig. 5a. Illustration of the operation of the method.

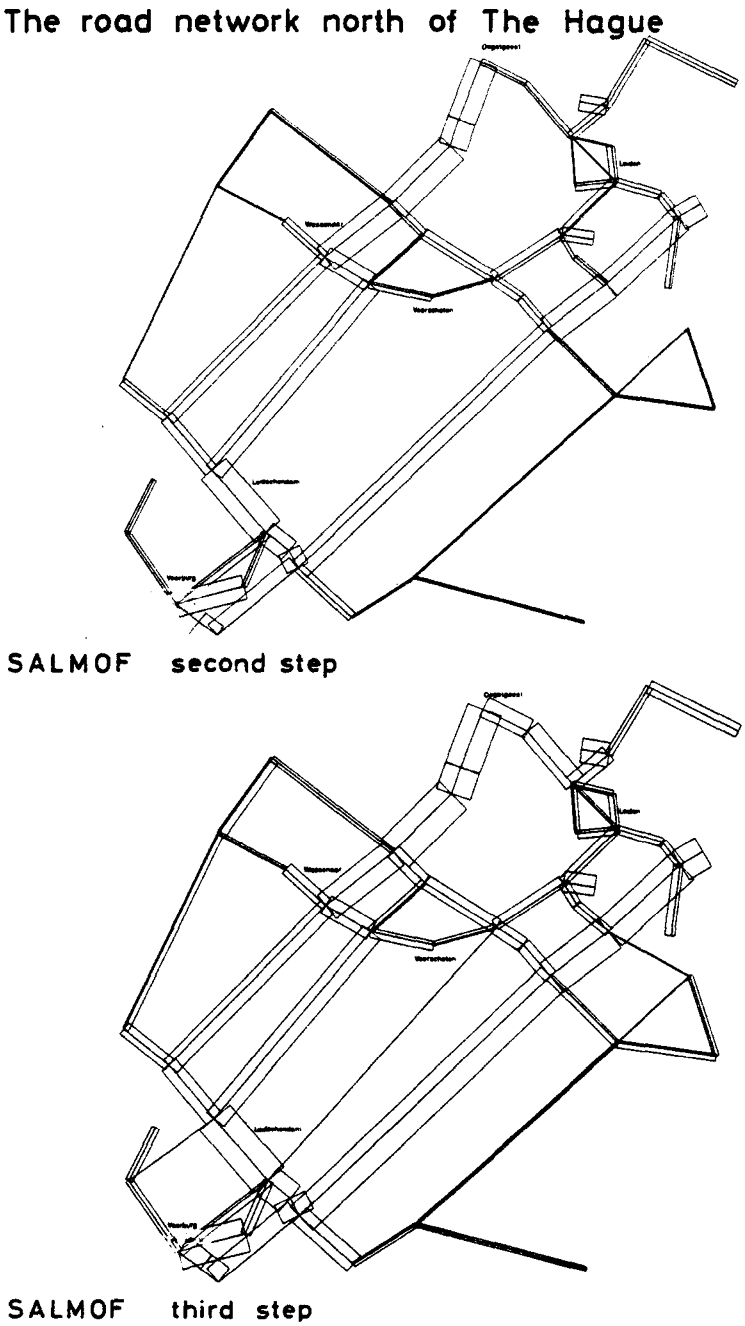


Fig. 5b. Illustration of the operation of the method.

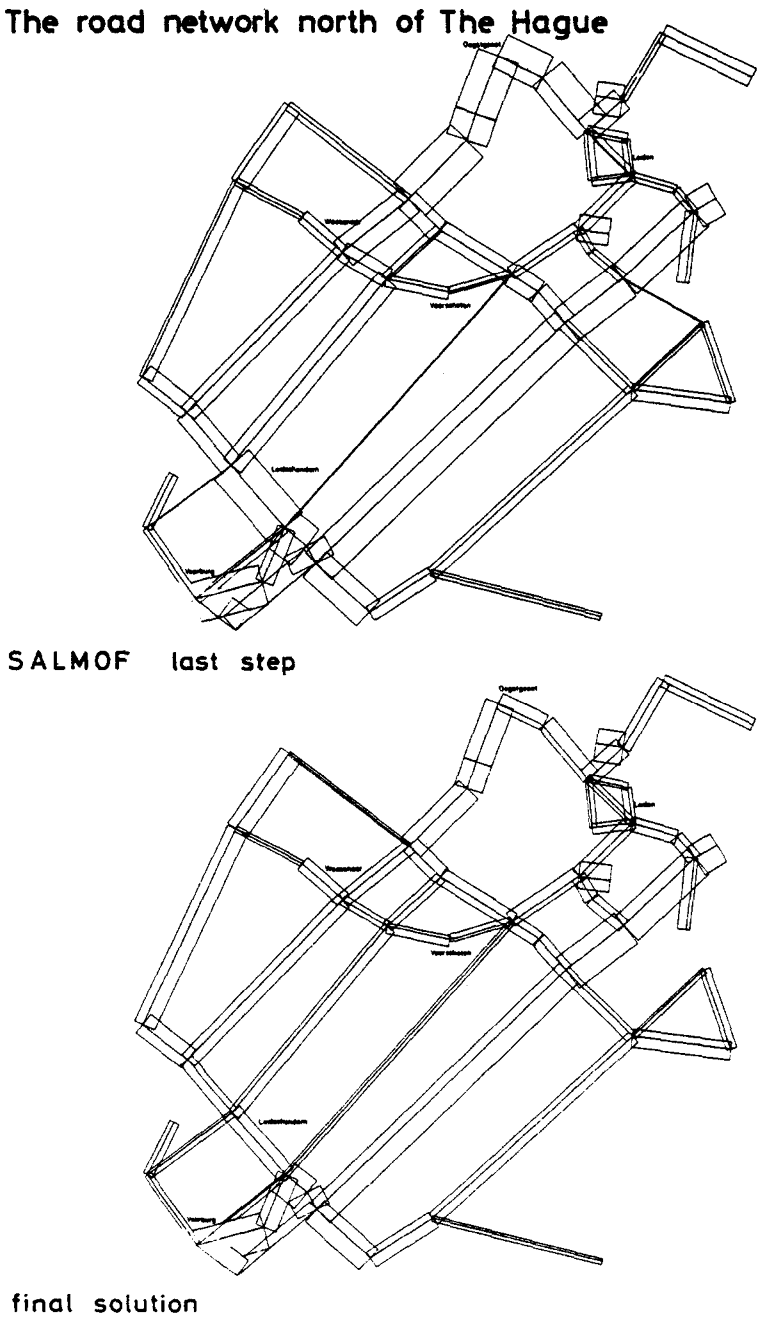


Fig. 5c. Illustration of the operation of the method.

lanes. Road 5 is expanded further. The flow on road 5 is now so heavy that an expansion above 2 × 4 lanes is "considered". This results in a small increase of the flow assigned to road 6 from the city, while road 4 also gets some traffic flow, though little.

At the fourth step roads 2 and 3 are so overloaded that they get no more traffic from the city. This traffic now takes the new road 1. Road 5 must now be expanded beyond 2 × 4 lanes. The investment costs involved with this are about as high as the cost for the construction of the new road 6. So we see a split of the flows between roads 5 and 6.

Because at the following descriptive assignment only the individual users cost are relevant we see at the final solution a shift of the flows on roads 1, 2 and 3 into the direction of road 1.* Furthermore we see that road 4 now gets a reasonable flow. Also, for this reason, the flow on road 5 decreases a little.

Results of the optimization

To get an idea of the usefulness of the optimization process proposed the results have been compared with the road network and flow pattern obtained with a heuristic procedure. This heuristic procedure consisted of a descriptive assignment to a network, in which all roads have their maximal dimension, followed by an optimal adjustment of the dimensions of the links to the traffic flows. Comparing the macroresults with each other the following could be seen:

(a) the optimization process gave a lower value for the objective function (about 15 per cent);
(b) in the optimization process hardly more kilometers had to be driven;
(c) in the optimization process there was less congestion;
(d) the amount of investments was considerably lower at the optimization process (12 per cent; 3·5 thousand million of Dutch guilders in absolute monetary terms).

At this it must be noted that the tripmatrix used contained so many trips, that for some parts of the country at both methods all road infrastructure possible had to be used. For tripmatrices containing less trips the differences would have been larger. Looking in detail at the networks and flow patterns the differences become clearer:

(a) generally the existing roads were better used in the optimization process;
(b) generally the construction and use of "expensive" roads were avoided in the optimization process;
(c) at some places in the optimization process clearly a certain choice was made between different roads, which were all used in the unlimited assignment;
(d) in some places the network obtained by the optimization process looked more coherent.

So for certain parts of the network a substantial difference in costs between the two networks appeared. It was also clear that in some places the optimal solution had not been found, for instance where two small tunnels were used instead of one larger one. That implied that a careful interpretation of the results would be necessary.

Concluding remarks

In this section we will go into the implications of the results of the Dutch Integral Transportation Study and especially of those of the road network optimization part of it.

Here six networks and their flow patterns were presented as the results of an optimization of a certain objective function, by use of certain decision variables and subject to certain constraints.

Many suppositions had to be made to arrive at this quantified problem formulation; this was especially the case at the defining of the constraints, which included decisive elements as the land-use pattern and travellers behaviour resulting in a certain tripmatrix.

Still it may be stated that the picture has been given, that will arise if the tripmakers respond to the circumstances in the same way as they do nowadays and if the government continues its policy with respect to the physical planning and the transportation planning.

The first question to be asked, was if this picture was desired or not. No official answer has been given, but generally the tendency seems to say "no". This "no" is argued to be caused by the inacceptable values of factors not taken into account, such as the situation inside towns and the damage to the environment, caused by the great number of highways and the heavy traffic flows on them. Also it seemed that the constraints were too rigid to allow for a desired solution; this last concerns especially the land-use and car ownership which formed important causes for the high amount of car kilometers. So new research is directed into two ways:

(a) how can the constraints be changed; that is, what alternatives are there for land use and travel behaviour, and how can one arrive at these?
(b) how can a more suitable objective function be stated, including factors not taken into account at the Dutch Integral Transportation Study?

* This shift is stimulated extra here, because the dimensions to and fro are set equal to the largest dimension before the descriptive assignment. This does not seem to be very sensible and hence this has not been done at the final computations of the Dutch Integral Transportation Study.

About the system to be developed there are two more-or-less opposite opinions:

(a) one extended optimization process, including the choice between the transport modes, the transport distribution and production and the land use;

(b) a combination of strategic search "by hand" on a high level and of automatic optimization on a lower level.

Of course combinations of these systems are possible.

Concluding, the Dutch Integral Transportation Study did not provide a ready-made transport infrastructure plan, but it provided a very important basis for discussion. It showed that principal decisions had to be taken and it has provided valuable instruments to be used in further research.

REFERENCES

Barbier M. (1966) Le future réseau de transports en région de Paris. *Cahiers de l'Institut d'Aménagement et d'Urbanisme de la Région Parisienne* **4-5** (4), 5–20.

Beckmann M. J., McGuire C. B. and Winsten C. B. (1956) *Studies in the Economics of Transportation.* Cowles Commission, Yale University Press, New Haven.

Bergendahl G. (1969) *Models for Investments in a Road Network.* Bonniers, Stockholm.

Bruynooghe M., Gibert A. and Sakarovitch M. (1968) Une méthode d'affectation du traffic, paper presented at the Fourth International Symposium on the Theory of Traffic Flow, Karlsruhe, June 1968.

Dafermos S. C. and Sparrow F. T. (1969) The traffic assignment problem for a general network. *Journal of Research of the National Bureau of Standards—B. Mathematical Sciences* **73B**, 91–118.

Hamerslag R. (1972) *Prognosemodel voor het Personenvervoer in Nederland.* Ph.D. Dissertation Technische Hogeschool Delft, Koninklijke Nederlandse Toeristenbond A.N.W.B., 's Gravenhage.

Haubrich G. Th. M. (1972) De optimalisering van het spoorwegnet in Nederland. Korte beschrijving van methode en resultaten. *Tijdschrift voor Vervoerswetenschap* **8.**

Highway Research Board (1966) *Highway Capacity Manual, 1965*, Washington DC.

Kuhn H. W. and Tucker A. W. (1951) Nonlinear programming. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (Edited by J. Neyman), University of California Press.

Manheim M. L., Ruiter E. R. and Bhatt K. U. (1968) Search and choice in transport systems planning: summary report, Research Report R68-40; M.I.T. Department of Civil Engineering, Cambridge, Mass.

Murchland J. D. (1969) Road Network Traffic Distribution in Equilibrium, paper presented at the conference: "Mathematical Methods in the Economic Science", Oberwolfach, Germany.

Nederlands Economisch Instituut (1972) *Rapport betreffende de Integrale Verkeers- en Vervoersstudie.* Staatsuitgeverij, 's Gravenhage.

Ochoa-Rosso F. and Silva A. (1968) Optimum project addition in urban transportation networks via descriptive assignment models, Research Report R68-44; M.I.T. Department of Civil Engineering, Cambridge, Mass.

Ridley T. M. (1968) An investment Policy to reduce the travel time in a transportation network. *Transpn Research* **2,** 409–424.

Steel M. A. (1965) Capacity restraint, a new technique. *Traffic Engng and Control.*

Steenbrink P. A. (1971) Optimalisering van de infrastruktuur. *Verkeerstechniek.* **22,** 251–262.

Steenbrink P. A. (1974) *Optimization of Transport Networks.* John Wiley, New York.

Timman R. (1966) *Optimaliseren van Funkties en Funktionalen.* Technische Hogeschool Delft, Onderafdeling der Wiskunde, Delft.

Tomlin J. A. (1969) Minimum-cost multicommodity network flows. *Ops Res.* **14,** 45–51.

Wardrop J. G. (1952) Some theoretical aspects of road traffic research. *Proceedings, Institute of Civil Engineers* **1,** 325–362.

Résumé—Dans la première partie de cet article on dresse le cadre de l'étude globale de transport hollandais et l'on insiste plus particulièrement sur la position du problème de définition du réseau routier. Dans la deuxième partie, le problème de l'optimisation du réseau de transport est décrit formellement et un algorithme envisagé pour résoudre ce problème est présenté. Le problème consiste à minimiser les coûts totaux de transport (coûts d'investissements et coûts de l'usager) d'une matrice de déplacement donnée, en choisissant les dimensions et les flux des maillons d'un réseau de transport. La méthode de résolution a été spécialement prévue pour surmonter les difficultés de calcul inhérentes aux réseaux importants. Elle fournit une solution raisonnable avec un temps de calcul relativement très court. A ce stade le programme est décomposé en plusieurs sous-programmes donnant la relation optimale entre la dimension et le flux de trafic pour chaque maillon, et en un programme principal affectant les flux au réseau de telle manière que la fonction objectif soit minimisée. Dans la troisième partie on décrit l'application de cette méthode d'optimisation au réseau routier hollandais.

Zusammenfassung—Im ersten Teil der Abhandlung werden der Gesamtrahmen der integrierten Verkehrsplanung für die Niederlande aufgezeigt und das Straßennetz in seinem derzeitigen Stand definiert. Der zweite Abschnitt enthält eine formale Darstellung des Netzoptimierungsproblems und eine Beschreibung des hierzu entwickelten Lösungsalgorithmus. Das Problem besteht darin, die Gesamtkosten (Investitionen und Benutzerkosten) bei einer gegebenen Fahrtenmatrix durch geeignete Wahl der Querschnitte und Belastungen der Netzstrecken zu minimieren. Das Lösungsverfahren wurde speziell mit der Absicht

entwickelt, die Schwierigkeiten bei der Berechnung großer Netze zu umgehen. Es liefert innerhalb relativ kurzer Rechenzeit ein durchaus annehmbares Ergebnis. Zu diesem Zweck wurde das Programm in eine Reihe von Unterprogrammen zerlegt, die zu einem optimalen Verhältnis zwischen Querschnitt und Belastung jeder Netzstrecke führen, und in einen übergeordneten Programmteil, der die Belastungen so über das Netz verteilt, daß die Zielfunktion minimiert wird. Im dritten Abschnitt schließlich wird die Anwendung des Verfahrens zur Optimierung des niederländischen Straßennetzes beschrieben.