ISSN 1822-8011 (print)
ISSN 1822-8038 (online)
INTELEKTINĖ EKONOMIKA
INTELLECTUAL ECONOMICS
2010, No. 1(7), p. 18–29

MODELLING AND OPTIMIZATION OF TRANSPORTATION COSTS

Gediminas DAVULIS

Mykolas Romeris University
Ateities str. 20, LT-08303 Vilnius, Lithuania
E-mail: geddav@mruni.lt

Linas ŠADŽIUS

Mykolas Romeris University
Ateities str. 20, LT-08303 Vilnius, Lithuania
E-mail: linassad@mruni.lt

Abstract. The article deals with ways of flow distribution in the transport network in order to minimize transportation costs. It describes the structure of transportation costs and principles of defining the dependence on the volumes of flow in pertinent transport modes, such as railways and road transport. It analyses classical models of flow distribution and presents a new approach to flow distribution in a transport network, based on flow optimization in individual contours or groups of contours of the network. The suggested approach is more rigorous than classical ones in mathematical terms and therefore avoids problems of heuristic nature that characterize classical approaches. The suggested approach has been tested in experimental calculations.

JEL classification: C02, D81.
Keywords: transportation costs, transport flows, distribution of transport flows.
Reikšminiai žodžiai: transportavimo išlaidos, transporto srautai, transporto srautų paskirstymas.

Introduction

Transportation is an economic sector that requires a lot of resources, the normal operation of which involves considerable labour, material and energy resources, while its modernization and further development require major investments. Therefore, it is very important to reduce these costs by employing technical, organizational and other means. One of the crucial means for reducing transportation costs is to distribute flows of cargo and passengers in the transport network in an efficient way. This, first of all, requires solving a problem which is known in the theory as the problem of optimal distribution of flows in the network. Its optimality criterion is the minimum of transportation costs. Based on the result, rational schemes of transportation of cargo and passengers can be developed. To solve this problem, it is not enough to make pilot assessments and primitive calculations. The complexity of these problems, as demonstrated by large volumes and non-linear relations, inevitably requires state-of-the-art approaches and use of modern computing equipment.

Presently, there is a wide range of approaches to optimizing transportation flows, where the transportation process is described in both linear and non-linear models (Bertsekas 1991; 1998; Gersht, Shulman, 1983; Kennington, Helgasson, 1980; Magnati, Wong, 1984; Minoux, 1989; Steenbrink, 1987; Golshtein, Sokolov 1995; Lifshich, 1986). In view of the importance and relevance of the problem, new results in this area are also important, especially if they can be used for practical purposes—solving problems of greater scope.

The article presents an approach to optimizing the distribution of transportation flows following a new approach, expanding the possibilities for solving such problems. The suggested approach is quite effective and can have a practical application.

1. General description of a transport system

Transport flow can be defined as the movement of vehicles carrying various types of cargo and passengers between different geographic destinations or within a certain region. The constituents or products of the transport flow are the components of the flow—flows of cargo of various types or passengers of various categories.

The transportation of passengers and cargo is carried out in a transport system consisting of fixed facilities—roads, railways, transport means (rolling-stock), using fixed facilities, and an organizational structure, ensuring the efficient usage of transport means and fixed facilities, and their interface. This system is usually described as a transport network consisting of nodes and arcs linking these nodes. The network nodes usually correspond to real transport junctions: stations, cities, reloading points, transition points (crossroads), and the arcs correspond to communications that connect the transport junctions—roads, railways, etc.

Each element of the transport network is characterized by one of its possible states and its load. The state of an element demonstrates its technical level. For example, a railway line can have some of the following states: a one-way line, a one-way line with automated blocking or a two-way line. Similarly, roads may be discriminated by the number of lanes. Obviously, each element of the network can be of only one possible state at a given point in time. The load of the network nodes may be defined as the volume and structure of products to be transported to or from a certain point. The load of the node is expressed by a vector, the number of coordinates of which equals to the number of products carried in the transport network. If any coordinate of this vector is equal to 0, this node is transitional with respect to the corresponding product; if it is positive, the node is a departure point (producer). The node is a destination point (user) if this coordinate is negative. Loads on the nodes determine transportation volumes in the transport network.

The load on a network arc is the sum of all constituents of the flow carried through it. The load of the arc is expressed by a vector, whose number of coordinates is also equal to the amount of carried products.

Flows in a transport network must meet the conditions of flow continuity, which relate the network topology and loading on its elements, namely, that the difference of product volumes transported to and from each node in the network must be equal to the load of that node by a respective product.

Transportation in a transport network is related to costs, which consist of direct costs, costs related to transportation of cargo or passengers (e.g., fuel costs), and maintenance costs of all components in the transport system. If management of certain volumes of transportation requires a reconstruction of individual elements in the network, it is also necessary to account for expenses for the development and modernization of the transport system. All expenses are comparable. Namely, we must analyse forced expenses, recalculated for a single time period. Expenses for transportation within a transport network are additive with respect to elements in the transport network, but are not additive with respect to flow constituents. If states of the network elements do not change, the main expenses are the maintenance costs of vehicles. The amount of these expenses per each element depends on the technical state and the load of this element. The dependence of maintenance costs on the load of a transport element is non-linear, because when the load approaches the margin of throughput for an element, expenses grow much faster than the load. This is not the only reason for this non-linear dependence. There are other natural reasons that differ according to the transport mode. If the state of the transport element does not change, this dependence is described by a convex function (Davulis, 1997). The exploitation costs may be reduced by reconstruction of the element to increase its throughput. In this case, however, the dependence of expenses and transportation volumes through an element becomes more complicated and is generally described as a non-convex function.

The distribution of transportation flows in a network defines the rational variation of passenger and goods flow in a transport network that meet the needs for the transportation of goods and passengers. Because there may be many such variations, there is need for criteria to measure them. The chosen criteria are the minimum individual and public expenses for transportation. Therefore, the problem of distribution of transportation flows may be described as an optimization problem.

2. Problem model for the distribution of transportation flows in a network

A model for the distribution of transportation flows assumes that flows of passengers and cargo carried in a transport network must be distributed in such a way that all needs of transport users would be satisfied at the lowest transportation cost. The problem of flow distribution can be analysed as a determined static problem, given invariable technical states of the network elements. Volumes of products of all types (types of cargo and categories of passengers), the number of which is defined as *n*, which must be sent from their departure points and received at their delivery points, the topology of the transportation network and the technical characteristics of its elements, determining the dependence of expenses for transportation via these network elements on the loads on these elements, i.e. on the volume of flows within them, are given. The topology of the transport network is described by an oriented graph with *M* nodes that correspond to real transportation points, and *L* arcs that correspond to the roads connecting the transportation points. The main constraints of the model of optimal flow distribution are flow continuity conditions that relate the needs of users, i.e. the transportation volumes, network topology, and flow volumes in the network arcs. If the conditions of network elements do not change, the quantity of transportation costs in each element can be described by a convex function on the load of this element in a rather precise way. Let's assume that the transportation costs are additive with respect to the network elements and are concentrated in the network arcs.

Under these assumptions, the product flow distribution in a transport network can be described by such a model:

$$\min_X f(X) = \sum_{l=1}^{L} f_l(x_l),$$
$$SX = B, \qquad (2)$$

here $f(X)$ is the function that expresses dependence of total expenses of transportation on the load of the network by a general flow, where $X = (x^{(1)},\ldots,x^{(n)})$; $x^{(j)} = (x_1^{(j)},\ldots,x_L^{(j)}$ is the vector of the load on the network by product *j*; $f_l(x_l)$ is the function of expenses (price) of a transport element (arc) that expresses the dependence of expenses on the load of the arc, where $x_l = (x_l^{(1)},\ldots,x_l^{(n)}$; $x_l^{(j)}$ is a variable expressing the load of arc *l* by product *j*; *S* is a quasi-diagonal $n \times n$ matrix, the main diagonals of which there are the node-arc incidence matrices of the given network, with zeros elsewhere; $B = (b^{(1)},\ldots,b^{(n)})$ is a general vector of the load of the network nodes that defines transportation volumes and their structure. $b^{(j)} = (b_1^{(j)},\ldots,b_M^{(j)})$ is the vector of the load of the network nodes by product *j*.

The system of equations (2) expresses conditions of flow continuity. If the transport network is described in an oriented graph, negative flows are also possible, indicating that products are transported in the direction opposite to the arc orientation. The general optimality criterion *f (X)* of the model is a convex function as a sum of the convex functions. Real functions of expenses, though convex, are most often non-smooth, which makes it difficult to solve the problem (1) – (2). The points at which the smoothness of the function is violated are called critical points. The paper deals with cases when the smoothness of the expense function is violated in a set of points, defined by linear equations called critical equations. Critical equations with only one variable $x_i^{(j)}$ with a non-zero coefficient are called main equations, and this variable is called the main variable. The arc is called critical if its load satisfies the critical equation of the expense function of this arc.

3. Modelling the expense function

The modelling of the expense functions is an independent and complex problem. The nature and specific expression of these functions depend on whether the system is normative or descriptive, on the transport mode and technical state of its elements. A specific expression of the expense function and the degree of its structure specification depend on both—the preciseness of the desired solution and on the availability of the necessary initial information and its accuracy. Usually, expenses are calculated per 1 km of road, and the total expenses for a road or its part are derived by multiplying expenses per 1 km by the length of the road. Expense functions calculated in such a way will differ only because of the level of the technical state of the road, and their number will depend on the number of the possible states of the elements in the transport system.

In normative systems, flows are planned based on the criterion of the economic-technical factors only, which reflect the transportation expenses of the transport system or the entire economy of the country. Freight railway transportation represents a classical example of a normative system.

As a rule, types of a railway line are distinguished by the number of roadways within it—whether it is one-way line, one-way line with a bypass, or a two-way line. Each variant is described by its own expense function. In all cases, expense functions are dependent on the flow or heaviness, expressed by the number of trains moving within a certain period of time in both directions. In the case of freight railway, expenses per one km of road are expressed by a general formula (Vasileva et al., 1981)

$$f(X^+, X^-) = f_d(X^+, X^-) + f_p(N_m). \qquad (3)$$

here X^+, X^- are the volumes of flows in each direction; $N_m = \max(N^+, N^-)$ is the heaviness of transport moving in one direction on a loaded trip, i.e. to the direction of higher heaviness; N^+, N^- are the numbers of freight trains to each direction.

Transition from flow volume to transport heaviness and vice versa is rather simple as those values are in linear relation

$$X = PN. \quad (4)$$

The first component of the expense function f_d is an expense component that corresponds to the expenses of a two-way line. The second component f_p represents additional expenses related to downtime in a one-way line with or without a bypass.

Maintenance costs in a two-way line are the sum of six expense components:

$$f_d(X^+, X^-) = \sum_i^6 g_i, \quad (5)$$

where $g_{i,}$ g_2 are the direct costs of transportation in both directions; g_3 represents direct costs while returning empty wagons; g_4 is expenses proportional to the time of transportation; g_5 is expenses proportional to the distance of cargoes transportation; g_6 is expenses proportional to freight volumes, expressed by flow volumes and traffic heaviness in the following way:

$$g_1 = a_1 X^+, \quad g_2 = a_2 X^-, \quad g_3 = a_3(N^+ - N^-)$$
$$g_i = a_{i1}N_m + a_{i2}X^+ + a_{i3}X^- + a(N^+ - N^-), \quad i = 4,5$$
$$g_6 = a_{61}(X^+ + X^-) + a_{62}N_m.$$

Coefficients $a_{1,}$ $a_{2,}$ a_3 depend on the length of the road part, type of locomotive (diesel locomotive or electric locomotive), the length of the station way, load on the axle of a loaded or empty wagon, price for diesel fuel, electric power, etc. The value of the coefficient a_3 depends on the returning direction of an empty wagon. The value of the coefficients in other expressions depends on the price of the locomotive hour, the price of the axle hour, expenses for work of teams servicing the locomotive, expenses per kilometre of the locomotive, kilometre of the axle, etc.

Additional expenses for a one-way railway line, related to a road intersection, are expressed as follows:

$$f_p(N_m) = \frac{(N_m + b_1)N_m^2}{b_2 - b_3N_m - b_4N_m^2} \quad (6)$$

Coefficients b_i in expression (6) are calculated based on the number of stops the trains make at intersections, downtime, and expenses for the train to start up and stop at stations, etc.

Expenses for a one-way line with a bypass with up to 70% road load are calculated in the same way as those for a two-way line, and above this load the additional expenses per 1 km are measured as follows:

$$f_p(N_m) = a(N_m - b)^2, \quad (7)$$

where coefficients a and b depend on the value of the road throughput and other technical parameters.

The methodology for calculation of all these coefficients is presented in the paper (Vasileva et al., 1981). This methodology could be used for railways of other countries, provided it is adjusted respectively, given present economic conditions.

In descriptive systems, interests of individuals or their groups and other subjective factors are of decisive importance for determining flows. Each transport participant follows an individual criterion of minimal expenses in those systems. In view of specific circumstances, the following criteria can be chosen: shortest distance, shortest transportation time, lowest transportation costs, most secure transportation, etc. The formation of a general criterion for a system is based on the hypothesis of collective behaviour, by assessing preferences of transport participants on the basis of observation data.

Road transport is a characteristic example of a descriptive system. Definite dependence of expenses on specific factors is difficult, but they may be considered adequate for real processes. In road transportation, expense functions are defined by the following components:

a) expenses dependent on the time of driving a vehicle.

b) expenses required for the operation of the vehicle—fuels costs, suspension, repairs, etc.

c) expenses related to accidents.

Scientists of the Dutch Institute of Economics suggested specific functions of expense types, expressed as gradual functions of one variable describing the degree of usage of road throughput, namely, traffic heaviness N in one lane (Steenbrink, 1987). The dependence of the expenses that are subjected to the time of driving of a vehicle on traffic heaviness in one lane is expressed as follows:

$$f_l = N\left[a_1 + a_2\left(\frac{N}{c}\right)^{a_4} + a_3\left(\frac{N}{c}\right)^{2a_4}\right], \quad 8)$$

where a_1, a_2, a_3, a_4 are coefficients derived from statistical data.

Functional dependence of expenses related to the operation of a vehicle and expenses related to accidents is expressed as follows:

$$f_a = N\left[b_1 + b_2\left(\frac{N}{c}\right)\right]^5 \quad (9)$$

The expense function which includes expenses necessary for transportation (e.g. fuel costs, payment for the carriers, etc.) and expenses for vehicle repairs depend on transportation volumes, and thus reflect variable costs. Meanwhile, expenses for the installation and improvement of transport infrastructure (roads, railways, buildings, warehouses, areas for passenger transport to stop or for reloading or storing of cargo, etc.), that is, investments required to improve a transport element have fixed costs.

Building of a new road for vehicles or the expansion of an existing one may throw the surrounding environment out of balance. Vehicle exhaust pollutes air, soil, water, and the noise has an adverse impact on people and fauna. This environmental damage depends on both, transport flow volume and traffic heaviness. However, it is difficult to quantify and measure it. Therefore, this damage is usually measured by introducing additional limitations on the problem, rather than by an expense function.

4. Classical solutions for the problem of flow distribution in a transport network

Classical methods for dealing with flow distribution problems for both normative and descriptive systems are divided into two stages. The first stage plans the transportation volumes and the structure between network nodes, in other words, defines the correspondences. All defined correspondences are entered into a table called a chess correspondence table or a transportation matrix. The second stage defines the load on the network elements by distributing the defined correspondences in the network by optimum routes in terms of the chosen criterion.

There are different principles for modelling correspondences. First of all, they differ depending on which system—normative or descriptive—is used for defining the correspondences. Descriptive systems demonstrate a wide range of principles for the modelling of correspondences. Here models may differ depending on transportation modes and their specific features.

In normative systems, volumes of products produced and consumed at specific points and their structure are usually known. In this case, correspondences are modelled by defining rational transport and economic relations from the viewpoint of an entire country. Classical or specific transport problems of linear programming are formulated and solved for this purpose.

If passengers are carried by railway, air or sea transport, statistical models of various types are most often used. Such models are also quite complex and account for many factors. In individual cases, such models are used for freight transportation, if the relationship between suppliers and users is short-term.

If passengers migrate between certain areas or regions, the so-called gravitation models are used to define correspondences between centres of these regions, which correspond to nodes of a transport network.

The so-called entropic models are mainly used for describing passenger correspondences in a city transport system. They model all types of travel, related to going to and from work, for cultural, routine and recreational purposes, and by using analogies between flows in transport and physical systems. These analogies are based on the fact that both systems contain a lot of uncontrolled interacting elements. In accordance with laws of thermodynamics, such systems aim at the condition of maximum entropy. Models of this type include optimization models with an objective function of an entropic type. This function usually takes into account preferences of the transportation participants by defining probabilities by means of a statistical survey p_{ab}, that a transportation participant travelling from point *a* will choose specifically point *b* as their destination point.

Once the transportation matrix has been created, loads on the network arcs by flows of cargo and passengers are defined. For this purpose, it is necessary to define an optimum route for each correspondence or its part. Parts of relevant correspondences the routes of which include this network arc are added together in each network arc. To deal with the problems of major scope and practical importance, methods based on the principle of rectification are used that can translate a non-linear optimization problem into sequence of linear problems. In every linear problem, correspondences or their parts are distributed into the shortest routes in terms of the optimum criterion, the "lengths" of their arcs being proportional to partial fluxions of the expense function. Although methods for the formation of correspondence routes may differ depending on the type of a transport system (normative or descriptive), as may the treatment of the optimum criterion, the principle model remains the same.

Optimum routes of correspondences are defined based on algorithms for finding the shortest route in a network. Though presently a number of such algorithms are known and of them are quite effective, this solution has been exhausted, because various known methods differ only in the ways of route formation. On the other hand, these algorithms contain a number of heuristic elements, the effective use of which requires relevant skills.

5. Suggested method for the optimization of flow distribution

In principle, the problem of flow distribution (1)–(2) can be dealt with by using any known non-linear programming method, provided that its scope is not overly great. However, problems of practical consequence are characterized by significant volumes, and thus possibilities for such solution are limited. Any known method will not be effective enough or even applicable if certain characteristics of the problems, such as large volumes or non-smooth objective function, are not taken into account. Therefore, dealing with practical problems of flow distribution requires special or modified methods. The paper suggests an original algorithm, created by the authors to solve the problem of flow distribution based on the modification of the idea of cyclic coordinate slope. The key principles of the suggested method have been presented in articles (Davulis, 1997; 1999).

Once the system of equations (2) is solved $N = n(L - M + 1)$ with respect to independent variables (it can be solved in a very simple way, by using graph means), the constraints of the problem may be described as such:

$$X = H\widetilde{X} + D, \qquad (10)$$

where H is a quasi-diagonal $n \times n$ matrix, on the main diagonal of which are network arc incidence matrices of respective products, and zeros elsewhere; $\widetilde{X}$ is an N-dimensional vector of independent variables; D is an L-dimensional vector of free terms of the system of equations, whose coordinates, corresponding to independent variables, are equal to zero. In this case, the system of equations (10) can be replaced by n independent system of equations that correspond to separate blocks of the matrix H:

$$x^{(j)} = H^{(j)}\widetilde{x}^{(j)} + d^{(j)}, \qquad j = 1,\ldots,n, \qquad (11)$$

where $H^{(j)}$ is the incidence matrix for arcs of the j-th network product, whose elements in the case of an oriented graph are -1; 0; 1.

Thus, the system of equations (10) allows the problem of conditional optimization (1) – (2) to be replaced by non-conditional optimization. As expense functions for individual elements of a transport network are usually non-smooth, "ditches" may occur in the total expense function. This may slow down the convergence of algorithms or generally they may become "stuck". The suggested algorithm provides for measures to eliminate or at least to mitigate such a phenomenon. Independent variables in the system of equations (10) always correspond to critical loads on the arcs. If, after any step of the optimization algorithm, this rule is violated, a procedure for the replacement of an independent variable in the system of equations (10) by a relevant basic variable must be performed. This procedure requires that the optimization be carried out in the direction of "a ditch" or close to it, provided that the point that corresponds to the flow distribution found is in this "ditch". This can improve the convergence of the algorithm.

The known theorems of graph theory (Christofides, 1976) relate basic and independent variables of the system of equations (11) that describe the distribution of an individual product in the network with the arcs of the graph that describe the network of that product in such a way. Every arc of the graph tree corresponds to basic variables in the system of equations, while arcs that do not belong to the tree—namely, free arcs—correspond to independent variables. This means that values of basic and independent variables in the system of equations j (11) are equal to loads in the relevant tree and free arcs by product j. Each column of the incidence matrix $H^{(j)}$ corresponds to a different free arc, and non-zero elements of this column correspond to tree arcs included in the contour formed by linking this free arc to the tree. Thus, each independent variable $x_i^{(j)}$ in the system of equations j (11) defines the contour unambiguously, defined by a free arc i in the network of product j.

By employing the indicated dependences between systems of equations describing distribution of individual products in the network and arcs of the graph describing the network, systems of equations (11) can be replaced by lists of network arcs, with tree and free arcs separated, for each product and array of the arc load made up of arc load vectors. In lists of network arcs, each arc corresponds to its code recorded in the load array. Lists of arc codes with their load arrays is the most compact way of writing systems of equations (11). However, to write systems of equations in this way, a special procedure is required: this procedure must distinguish codes of any contour arcs from the list of the network arcs. This procedure has been described in a previous paper (Davulis, 1997).

Solution to the problem can be divided into three stages: 1) initial flow distribution; 2) initial optimization; 3) main optimization.

Initial flow distribution requires transformation of the system of equations (2) into system (11) by using graph means. This procedure has been described in an earlier paper (Davulis, 1997). Upon completion of this procedure, we will have a tree with arc loads that meet the conditions of flow continuity; loads of all other arcs are equal to zero. In other words, flow corresponds to basic solutions of systems of equations (11).

If the initial tree is selected randomly, corresponding initial flow distribution may be far from optimum. Therefore, it makes sense to select a more proper tree with better flow distribution in terms of the optimum criterion. Initial optimization (iterative procedure) may be used to minimize the total functions of arcs of each contour in turn in the set of zero loads of these contour arcs. If the load on the tree arc that belongs to the contour to be optimized in any iteration becomes equal to zero, codes of this arc and the free arc which formed this contour in the lists of the tree and free arcs are replaced by each other by using a relevant procedure. This procedure is repeated in cycles for each contour in the network and is completed when the tree does not change after completion of iterations $L-M$ successively. In the case of convex expense functions, this initial optimization creates flow distribution in the network which is closer to the optimum, and this leads to reduction of the scope of calculations.

The main optimization is basically an algorithm of contour optimization. The algorithm consists of two stages repeated in cycles. The first stage of the algorithm is an iterative procedure, in each iteration of which the distribution of one product in the network is optimized, given unchanged distribution of all other products. In general, the algorithm uses individual lists of the network arcs for each product. In the iteration of the algorithm, in which the distribution of product *j* is optimized, the total expense function of the arcs of each contour is minimized in turn, i.e. the following problem is solved for each contour:

$$\min_{a \le \Delta x_k^{(j)} \le b} F_k^{(j)}(\Delta x_k^{(j)} = \sum_{i \in \mathbf{I}_k^j} f_i(x_i \parallel \bar{x}_i^{(j)} + e_{ik}^{(j)} \Delta x_k^{(j)}), \quad (12)$$

where *j* is the number of the product to be distributed in the network in the given iteration; *k* is the number of the free arc that formed the contour; $\bar{x}_i^{(j)}$ is the initial value of the variable $x_i^{(j)}$; $\mathbf{I}_k^j$ is the set of numbers of the graph arcs belonging to the network contour of product *j* defined by arc *k*; $e_{ik}^{(j)}$ is the coefficient with value of 1, provided that directions of arc *i* and free arc *k* concur in the contour, and -1 in the opposite case. This coefficient is equal to the element of incidence matrix $H^{(j)}$ in the line of variable $x_i^{(j)}$ and column of variables $x_k^{(j)}$.

The interval of optimization [*a, b*] is the smallest interval of the changes in the load on the free arc by product *j*, which reaches zero load of any contour by this product. A procedure of single-dimensional search is used for the solution of problem (12); this procedure is a modification of the golden section search. Loads on the contour arcs by product *j* are recalculated respectively if the optimal value of the change in the flow of product *j* and arc *k*, if it became critical and is replaced in the tree arc list by a free arc that defined this contour. This procedure of tree replacement corresponding to the replacement of the independent variable in the system of equations (10) by a basic variable means that we will be able to minimize the set of points in which smoothness of the expense function is violated.

The optimization procedure (12) is repeated in cycles for each network contour of product *j* until the condition for terminating of the algorithm is satisfied, namely, change in values of the total expense function of each contour is lower than the set value $\varepsilon 1$. Iterations of the algorithm are repeated in cycles for each product until a distribution of the general flow is obtained that meets the conditions for terminating the algorithm. If the flow distribution obtained satisfies critical equations of one or two arc expense functions of a general form—which include at least two coordinates of the arc load vector with non-zero coefficients—there is no guarantee that the local minimum is found in the required preciseness, though the conditions for terminating the algorithm are met. Such a situation may mean that the flow distribution derived corresponds to the non-smoothness point of the price function, and optimisation should be continued in the set of points defined by these critical lines and conditions (11), carrying out the second stage of the algorithm.

Critical equations of a general form that meet the flow derived are expressed as a system of equations:

$$y_p = L_{i_p}(x_{i_p}) = 0, \qquad p = 1, \ldots, P, \qquad (13)$$

where $L_{i_p}(x_{i_p})$ is a linear function; *P* is the number of critical equations. Variables x_{i_p} of critical equations (13) are replaced by independent variables that express loads of free arcs by using dependencies:

$$x_i^{(j)} = \hat{x}_i^{(j)} + \sum_{k \in \mathbf{K}_i^j} e_{ik}^{(j)} x_k^{(j)}, \qquad i \in \bigcup_{k \in \mathbf{K}_i^j} \mathbf{I}_k^j, \qquad j \in \mathbf{J}_{\mathbf{i}}, \quad (14)$$

where $\hat{x}_i^{(j)}$ is the initial value of the variable that expresses the load of the arc *i* by product *j*; $\mathbf{K}_i^j$ is the set of numbers of free arcs to which contours in the network of product *j* the critical arc belongs; $\mathbf{J}_i$ is the set of numbers of the coordinates of the load vector of the critical arc *i* which are included in the critical equation of this arc with non-zero coefficients.

Subsequently, the system of equations (13) is solved in terms of independent variables *N* – *P* and thus we get a new system of equations:

$$\overline{X} = \overline{H}Z, \qquad (15)$$

where the coordinates of vector $\overline{X}$ of the basic variables—to be called non-main variables—are a part of the coordinates of vector $\widetilde{X}$, and the coordinates of vector *Z* of independent variables are the remaining coordinates of vector $\widetilde{X}$. System of equations (15) is recorded in the form of a table, to be called a table of non-main variables. With variation of the value of independent variable z_k, the values of those non-main variables that are in the lines of non-zero elements of column *k* will also vary. Their values change in order to meet critical equations (13). With a change in the value of a non-main variable $x_i^{(j)}$, the loads of all arcs in the contour formed in the network of product *j* by arc *i* will change, respectively. Thus, column *k* of the table of non-main variables defines a group of contours whose loads of the arcs by relevant products will change, given changes in the value of independent variable z_k. This group of contours are called related contours.

The second stage of the algorithm is a procedure of iterative one-dimensional optimization, which is repeated in cycles for each independent variable z_k until the condition for terminating the algorithm is met. It employs the same optimization procedure as in the first stage of the algorithm, only in this case the summary expense function of all related contour arcs is minimized. The optimization procedure in the case of related contours is written as follows:

$$\min_{a \le \Delta z_k \le b} F_k(\Delta z_k) = \sum_{l \in \mathbf{L}^k} f_l(x_l(\Delta z_k)), \qquad (16)$$

where [*a*, *b*] is the minimum range of the independent variable z_k which reaches zero load of any arc of the related contours by a product the flow of which changes in the curve, given changes in the independent variable z_k; $\mathbf{L}^k$ is the set of curves of all related contours.

Any arc of the related contour group belongs to one or several contours of this group, defined in the network of different or the same product. Therefore, such an arc may be subject to changes in the load by either one or several products, while load changes have a linear dependence on changes Δz_k in the independent variable z_k. Any change in the load on arc *l* of the related contours by separate products are expressed in a general formula:

$$\Delta x_l^{(j)} = \sum_{s \in \mathbf{S}_l^j} e_{sk}^{(j)} \overline{h}_{sk}^{(j)} \Delta z_k, \qquad j \in \overline{\mathbf{J}}_l^k, \qquad l \in \mathbf{L}^k, \quad (17)$$

here $\mathbf{S}_l^j$ is the set of numbers of the free arcs to which contours formed in the network of products *j* arc *l* belong; $\overline{\mathbf{J}}_l^j$ is the set of numbers of those products whose flows change in arc *l* subject to changes in variable z_k.

If some arcs of the related contours become critical after the second stage of the algorithm, system of equations (13) is supplemented by new critical equations, and the aforementioned procedure is followed. The second stage of the algorithm is completed when no new critical arcs occur after routine iteration or when after each iteration, consecutively decreasing of the number of the independent variable z_k, the last iteration determines an optimal change in the single remaining independent variable z_1. This is followed by a return to the first stage of the algorithm and the described procedure is repeated until the flow does not change after the first or second stage of the algorithm. This distribution represents a local solution to the problem derived with a given precision.

The second stage of the algorithm may be necessary if critical sets are described by critical equations of a general form. When critical equations are only the main ones, the first stage of the algorithm is sufficient, as the procedure for altering the graph structure by changing the codes of free and tree arcs in the arc lists guarantees convergence to the solution of the problem determining it with given precision.

The suggested algorithm demonstrates a number of advantages compared to traditional approaches. It is rigorous in mathematical terms; therefore, it presents no heuristic problems, unavoidable in classical approaches. The quality of flow distribution in terms of the optimum criterion depends on the volume of the part of the correspondence to be distributed in the network, accuracy of calculations and frequency of recalculation of differential expenses. The relationship between these parameters is often complicated. The best relationship between these parameters is usually defined experimentally. On the other hand, classical approaches are not applicable in the case of non-smooth expense functions, particularly, when "ditches" in the values of these functions appear. As a result, the convergence of classical approaches slows down or may become "stuck" far from the point that would constitute a solution. To avoid this situation, the algorithm must be supplemented by special complex procedures based on the approximation of the expense function in its field of non-smoothness by a smooth function, accelerating its convergence.

The suggested algorithm does not require the creation of a transformation matrix, and ensures a solution of the required precision. Meanwhile, the classical model of solution in two stages: definition and distribution of correspondences may not always ensure a solution close to the optimum. The algorithm employs an efficient procedure of a one-dimensional optimization—namely, a modified golden section procedure. It is therefore sufficiently effective from the point of view of these calculations, as also demonstrated by practical calculations. The most important advantage of this algorithm lies in that it is applicable in the cases of non-smooth functions as well.

The algorithm has been tested in the experimental calculations, solving flow distribution problems in real transport networks described in other publications. A number of classical approaches were compared to the proposed model in the calculation of the distribution of goods flows in a network consisting of 43 transport points and 49 railway lines linking them, of which 27 are one-way lines, and are 22 two-way lines. As indicated in the paper (Vasileva et al.,1981) this network is part of a real railway network. This paper presents the calculation results obtained by classical approaches by using simplified transportation expense functions, as well as dependence of the values obtained by optimum criterion (expense function) on the number of iterations.

The transportation flow distribution was calculated by using the same transportation expense functions in the same network; it was calculated by a using simplified version of the suggested contour optimization algorithm wherein only in the first stage of the algorithm was used. The calculation results and their comparison with results obtained by traditional approaches have been presented in a previous article (Davulis, 2000). After performing 50 iterations, the contour optimization algorithm produced a result much similar to the optimum distribution of flows obtained by classical approaches after the same number of iterations. The value of the optimum criterion (transportation costs) obtained by the suggested approach was by 2-5 percent lower than that obtained by traditional approaches, and about 3 percent higher if compared to the result obtained by the consecutive distribution approach, which employs special measures accelerating the convergence of the algorithm. These results prove that the suggested contour optimization algorithm is appropriate for calculating the distribution of transportation flows in a network. It hardly makes sense to question which of the algorithms analysed is the best, unless it were possible to analyse which algorithm is more appropriate given specific conditions. However, this requires a lot of calculations with different networks and different parameters. Thus, practice should produce an answer to this question.

Results of the calculations highlighted one more advantage of the suggested algorithm. Algorithms based on lining principles are irregular given an increase in the number of iterations. They do not ensure a consecutive decrease in the optimum criterion, as linear approximation is rather approximate. Meanwhile, the suggested algorithm is regular, which allows to avoid unnecessary iterations of the algorithm.

Conclusions

Successful distribution of transportation flows in a network leads to a reduction in transportation costs. Therefore, studies in this field are important from both, the theoretical and the practical points of view.

The scope of practical problems in transportation flow distribution is formidable—it covers large networks, a considerable variety of cargo and categories of passengers. This is why the path followed by classical approaches in solving these problems is absolutely natural: to replace a non-linear problem of a large scope by a succession of linear problems, as modern approaches can solve problems of transportation flow distribution of a large scope only in a linear manner. However, problems based on the lining principle have some essential drawbacks. First of all, they require initial formation of correspondences which are afterwards distributed in the transport network for routes that are the shortest in terms of the optimum criterion. Secondly, these algorithms include a number of heuristic elements—their effectiveness depends on the proper selection of succession parameters, and this requires certain skills. These algorithms are also irregular. They are not applicable in case of non-smooth optimum criteria.

Classical approaches based on the lining principle have been seen as having primary and unique significance up to now. As an alternative, the article suggests an original approach based on optimization of transportation flows in the individual contours or in their groups. The suggested approach of contour optimization has obvious advantages: there is no need for advance formation of correspondences, it is rigorous in mathematical terms; it is regular and appropriate in cases of non-smooth transportation cost functions. A special way to encode the transport network and the use of graph theory measures allows for the provision of all information in the most compact way, making this algorithm useful in solving problems of large scope—those with practical importance. Experimental calculations proved this algorithm to be absolutely appropriate in solving problems of transportation flow distribution, and the results derived were not worse or were close to the results derived by traditional approaches.

References

1. Bertsekas D.P. (1991). Linear network optimization: algorithms and codes. Cambridge, Mass: MIT Press.
2. Bertsekas D.P. (1998). Network optimization: continuous and discrete methods. Belmont, Mass: Athena Scientific.
3. Christofides N. (1976). Graph theory. An algorithmic approach. Academic Press, New York.
4. Davulis G. (1997). Optimization of transport flows (in Lithuanian) // Transportas, 1997, 1(14). Vilnius: Technika, p. 18–34.
5. Davulis G. (1999). A method of contour optimization for transport flow distribution // Statistics in Tranzition, V.4, No.2, p. 2001-212.
6. Davulis G. (2000). Methods for optimizing transport flows and their experimental comparison (in Lithuanian) // Transportas, XV t., Nr. 2 Vilnius: Technika, p. 60–71.
7. Gersht A., Shulman A. (1983). A new algorithm for the minimum cost multicommodity flow problem // Proc. 26 Conf. Die and Control. Los Angeles, p. 748–758.
8. Kennington J.L. and Helgasson A.V. (1980). Algorithms for network Programming. New York, John Willey and Sons.
9. Magnati T.L., Wong R.T. (1984). Network design and transportation planning models and algorithms // Transportation science, v. 18, No. 1 , p. 3–55.
10. Minoux M. (1989). Network synthesis and optimum network design problems: models, solution methods and applications // Networks, v. 19, No. 2, p. 313–360.
11. Steenbrink P.A. (1987). Optimization of Transport Networks. London – New York – Sydney, Toronto.
12. Vasileva E.M, Levit B.J, Lifshich V.H. (1981). Non-linean transport problems on networks (in Russian). M.: Финансы и статистика.
13. Golshtein E.G., Sokolov H.A. (1995). The new method for solving multicommodity transport problem (in Russian) // Экономика и мат. методы, т. 31, вып. 2, с. 128–136.
14. Lifshich V. H (1986). The systemic analysis of economic processess in transport (in Russian). M.: Транспорт.

SUPPLEMENTS

Supplement 1

Oriented graph describing the transport network and the corresponding node-arc incidence matrix

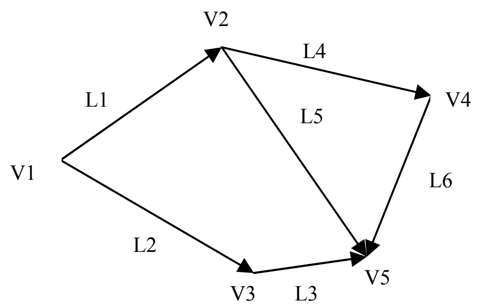


Figure 1. Oriented graph describing the transport network consisting of 5 (V1, V2, ...,V5) transportation points and six (L1, L2, ..., L6) communication lines linking transport networks.

A transport network can be illustrated graphically as an oriented or non-oriented graph with a topology described by its matrix of incidences. Each line of the incidence matrix corresponds to a graph arc, and each column corresponds to an individual node in the graph. Each element of the incidence matrix of the non-oriented graph may have values of either 1 or 0 reflecting whether the relevant arc and the node are linked or not. In an oriented graph (graph with arcs that have a defined direction), an element of the incidence matrix equals -1 if the relevant arc goes towards the node, and equals +1 if the arc goes away from it. Another graph with a separate incidence matrix may correspond to each constituent of the flow.

	V1	V2	V3	V4	V5
L1	1	-1	0	0	0
L2	1	0	-1	0	0
L3	0	0	1	0	-1
L4	0	1	0	-1	0
L5	0	1	0	0	-1
L6	0	0	0	1	-1

Figure 2. Incidence matrix that corresponds to the oriented graph shown in Figure 1.

Supplement 2

Graph with separated tree and free arcs, and the corresponding incidence matrix for the arcs.

Separation of the network arcs into tree and free arcs is graphically shown in Figure 3; the incidence matrix for arcs which corresponds to this separation is shown in Figure 4. The network tree consists of all arcs in the network which do not form contours in the network. The selection of the tree is not unambiguous.

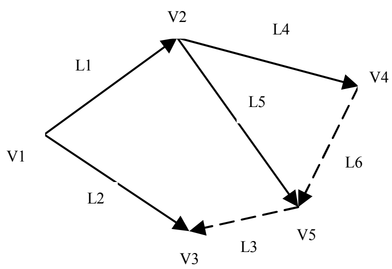


Figure. 3. Graph reflecting the transport network with free (L3 and L6) and tree (L1, L2, L4, L5) arcs separated.

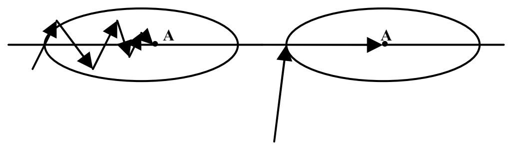
	L3	L6
L1	-1	0
L2	1	0
L4	0	1
L5	-1	-1

Figure 4. Incidence matrix for arcs that corresponds to the oriented graph shown in Figure 1

Elements of the incidence matrix that correspond to the tree arcs equal -1 if their directions in the contour are opposite to the direction of the free arc that defines this contour; they equal 1 if these directions concur, and they equal 0 if the contour does not include tree arcs. If flow in the free arc (e.g., L6) increases by value ΔX, then flow in the tree arc L4, whose concurs with the direction of the free arc will increase by the same value, and in tree arc L5, whose direction is opposite to the direction of the free arc L6, flow will decrease by the same value.

Supplement 3

Graphical comparison of the convergence of classical and suggested algorithms towards solution A



a) classical algorithm b) suggested algorithm

Figure 5 The nature of convergence of the classical (a) and suggested (b) algorithms towards the solution in the "ditch".

TRANSPORTAVIMO IŠLAIDŲ MODELIAVIMAS IR OPTIMIZAVIMAS

Gediminas DAVULIS, Linas ŠADŽIUS
Mykolo Romerio universitetas, Lietuva

Santrauka. Straipsnyje nagrinėjami srautų paskirstymo būdai transporto tinkle siekiant sumažinti transportavimo išlaidas. Aptariamos transporto srautų formavimosi ypatybės atskirose transporto sistemose, t. y. normatyvinėse, kai srautai formuojami vadovaujantis vienu ekonominio-techninio turinio kriterijumi, rodančiu visuomenės išlaidas transportavimui, ir deskriptyvinėse, kai kiekvienas važiuojantysis vadovaujasi savo individualiu išlaidų minimumo kriterijumi, sistemose. Aprašoma transportavimo išlaidų struktūra bei jų priklausomybės nuo srautų dydžio nustatymo principai tokių Lietuvai svarbių rūšių kaip geležinkelių ir automobilių transporte. Nagrinėjami klasikiniai srautų paskirstymo modeliai bei srautų problemų sprendimo būdai. Pateikiamas naujas srautų paskirstymo būdas transporto tinkle, pagrįstas srautų optimizavimu atskirose tinklo kontūruose arba jų grupėse. Siūlomas metodas matematiniu požiūriu yra griežtesnis negu klasikiniai ir todėl čia išvengiama euristinio pobūdžio problemų, būdingų klasikiniams metodams. Eksperimentiniai skaičiavimai parodė, jog siūlomas metodas gali būti taikomas ir praktinę reikšmę turintiem uždaviniams spręsti.

Gediminas Davulis – Professor, Dr., Department of Economics, Mykolas Romeris University, Vilnius Lithuania. Research areas: modelling and investigation of macroeconomic processes, public economics, management of processes in economy and transport, application of optimization methods in economics. Author of more than 80 scientific publications.

Gediminas Davulis – Mykolo Romerio universiteto Ekonomikos ir finansų valdymo fakulteto Ekonomikos katedros profesorius. Paskelbė daugiau kaip 80 mokslo darbų. Pagrindinės tyrimų kryptys – makroekonominių procesų modeliavimas ir tyrimas, viešojo sektoriaus ekonomika, ekonomikos ir transporto vadyba, optimizacinių metodų taikymas ekonomikoje.

Linas Šadžius – Assoc. Professor at the Department of Investment and Banking, Faculty of Economics and Finance Management, Mykolas Romeris University, Vilnius, Lithuania. Research areas: analysis of macroeconomic processes and banking activities.

Linas Šadžius – Mykolo Romerio universiteto Ekonomikos ir finansų valdymo fakulteto Investicijų ir bankininkystės katedros docentas. Tyrimų kryptys – makroekonominių procesų bei bankų veiklos analizė.