On the history of the transportation and maximum flow problems

Alexander Schrijver

Abstract. We review two papers that are of historical interest for combinatorial optimization: an article of A.N. Tolstoi from 1930, in which the transportation problem is studied, and a negative cycle criterion is developed and applied to solve a (for that time) large-scale (10 × 68) transportation problem to optimality; and an, until recently secret, RAND report of T.E. Harris and F.S. Ross from 1955, that Ford and Fulkerson mention as motivation to study the maximum flow problem. The papers have in common that they both apply their methods to the Soviet railway network.

1. Transportation

The transportation problem and cycle cancelling methods are classical in optimization. The usual attributions are to the 1940’s and later\(^{2}\). However, as early as 1930, A.N. Tolstoi \cite{1} published, in a book on transportation planning issued by the National Commissariat of Transportation of the Soviet Union, an article called *Methods of finding the minimal total kilometrage in cargo-transportation planning in space*, in which he studied the transportation problem and described a number of solution approaches, including the, now well-known, idea that an optimum solution does not have any negative-cost cycle in its residual graph\(^{3}\). He might have been the first to observe that the cycle condition is necessary for optimality. Moreover, he assumed, but did not explicitly state or prove, the fact that checking the cycle condition is also sufficient for optimality.

Tolstoi illuminated his approach by applications to the transportation of salt, cement, and other cargo between sources and destinations along the railway network of the Soviet Union. In particular, a, for that time large-scale, instance of the transportation problem was solved to optimality.

We briefly review the article. Tolstoi first considered the transportation problem for the case where there are only two sources. He observed that in that case one can order the destinations by the difference between the distances to the two sources. Then one source can provide the destinations starting from the beginning of the list, until the supply of that source has been used up. The other source supplies the remaining demands. Tolstoi observed that the list is independent of the supplies and demands, and hence it

is applicable for the whole life-time of factories, or sources of production. Using this table, one can immediately compose an optimal transportation plan every year, given quantities of output produced by these two factories and demands of the destinations.

---

\(^{1}\) CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands, and Department of Mathematics, University of Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands.

\(^{2}\) The transportation problem was formulated by Hitchcock \cite{1}, and a cycle criterion for optimality was considered by Kantorovich \cite{1} (Kantorovich and Gavurin \cite{1}), Koopmans \cite{1} (Koopmans and Reiter \cite{1}), Robinson \cite{1,1}, Gallai \cite{1,1}, Lur’e \cite{1}, Fulkerson \cite{1}, and Klein \cite{1}.

\(^{3}\) Later, Tolstoi described similar results in an article entitled *Methods of removing irrational transportations in planning* \cite{1}, in the September 1939 issue of *Sotsialisticheskii Transport*.

\(^{4}\) The *residual graph* has arcs from each source to each destination, and moreover an arc from a destination to a source if the transport on that connection is positive; the cost of the ‘backward’ arc is the negative of the cost of the ‘forward’ arc.
Figure 1
Figure from Tolstoi [1930] to illustrate a negative cycle.

Next, Tolstoi studied the transportation problem in the case when all sources and destinations are along one circular railway line (cf. Figure 1), in which case the optimum solution is readily obtained by considering the difference of two sums of costs. He called this phenomenon circle dependency.

Finally, Tolstoi combined the two ideas into a heuristic to solve a concrete transportation problem coming from cargo transportation along the Soviet railway network. The problem has 10 sources and 68 destinations, and 155 links between sources and destinations (all other distances are taken to be infinite), as given in the following table.
<table>
<thead>
<tr>
<th></th>
<th>Arkhangelsk</th>
<th>Vologda</th>
<th>Murom</th>
<th>Balaklonsk</th>
<th>Derzhinik</th>
<th>Kuhin</th>
<th>Sverdlovsk</th>
<th>Artemovsk</th>
<th>Derzhik</th>
<th>Vologda-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>supply</td>
<td>5</td>
<td>11.5</td>
<td>8.5</td>
<td>12</td>
<td>100</td>
<td>12</td>
<td>15</td>
<td>5</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>demand</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table of distances (in kilometers) between sources and destinations, and of supplies and demands (in kilotons). (*Kolstof did not give any distance for Kasinov. We have inserted a distance 0 to Murom, since from ‘Kolstof’s solution it appears that Kasinov is connected only to Murom, by a waterway.*)
Tolstoi’s heuristic also makes use of insight in the geography of the Soviet Union. He goes along all sources (starting with the most remote sources), where, for each source \( X \), he lists those destinations for which \( X \) is the closest source or the second closest source. Based on the difference of the distances to the closest and second closest sources, he assigns cargo from \( X \) to the destinations, until the supply of \( X \) has been used up. (This obviously is equivalent to considering cycles of length 4.) In case Tolstoi foresees a negative-cost cycle in the residual graph, he deviates from this rule to avoid such a cycle. No backtracking occurs.

In the following quotation, Tolstoi considers the cycles Dzerzhinsk-Rostov-Yaroslavl’-Leningrad-Artémovsk-Moscow-Dzerzhinsk and Dzerzhinsk-Nerekhta-Yaroslavl’-Leningrad-Artémovsk-Moscow-Dzerzhinsk. It is the sixth step in his method, after the transports from the factories in Ilets, Sverdlovsk, Kishert, Balakhonikha, and Murom have been set:

6. The Dzerzhinsk factory produces 100,000 tons. It can forward its production only in the Northeastern direction, where it sets its boundaries in interdependency with the Yaroslavl’ and Artémovsk (or Dekonskaya) factories.

<table>
<thead>
<tr>
<th>From Dzerzhinsk</th>
<th>From Yaroslavl’</th>
<th>Difference to Dzerzhinsk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berendeëvo</td>
<td>430 km</td>
<td>135 km</td>
</tr>
<tr>
<td>Nerekhta</td>
<td>349 km</td>
<td>50 km</td>
</tr>
<tr>
<td>Rostov</td>
<td>434 km</td>
<td>56 km</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>From Dzerzhinsk</th>
<th>From Artemovsk</th>
<th>Difference to Dzerzhinsk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aleksandrov</td>
<td>397 km</td>
<td>1,180 km</td>
</tr>
<tr>
<td>Moscow</td>
<td>405 km</td>
<td>1,030 km</td>
</tr>
</tbody>
</table>

The method of differences does not help to determine the boundary between the Dzerzhinsk and Yaroslavl’ factories. Only the circle dependency, specified to be an interdependency between the Dzerzhinsk, Yaroslavl’ and Artémovsk factories, enables us to exactly determine how far the production of the Dzerzhinsk factory should be advanced in the Yaroslavl’ direction.

Suppose we attach point Rostov to the Dzerzhinsk factory; then, by the circle dependency, we get:

<table>
<thead>
<tr>
<th>Dzerzhinsk-Rostov</th>
<th>454 km</th>
<th>-398 km</th>
<th>Nerekhta</th>
<th>349 km</th>
<th>-299 km</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yaroslavl’-</td>
<td>56</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yaroslavl’-Lenigrad</td>
<td>709</td>
<td>+938</td>
<td>These points remain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Artémovsk-</td>
<td>1,667</td>
<td>unchanged because only the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Artémovsk-Moscow</td>
<td>1,030</td>
<td>-625</td>
<td>quantity of production sent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dzerzhinsk-</td>
<td>405</td>
<td></td>
<td>by each factory changes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Total            | -65 km | +34 km |

Therefore, the attachment of Rostov to the Dzerzhinsk factory causes over-run in 65 km, and only Nerekhta gives a positive sum of differences and hence it is the last point supplied by the Dzerzhinsk factory in this direction.

As a result, the following points are attached to the Dzerzhinsk factory:
N. Novgorod  25,000 tons
Ivanova   6,000 ..  
Nerechta  5,000 .. 
Aleksandrov 4,000 .. 
Berezhnoe 10,000 .. 
Likino    15,000 .. 
Moscow    35,000 ..  (remainder of factory’s production)

Total 100,000 tons

After 10 steps, when the transports from all 10 factories have been set, Tolstoi “verifies” the solution by considering a number of cycles in the network, and he concludes that his solution is optimum:

Thus, by use of successive applications of the method of differences, followed by a verification of the results by the circle dependency, we managed to compose the transportation plan which results in the minimum total kilometrage.

The objective value of Tolstoi’s solution is 395,052 kiloton-kilometers. Solving the problem with modern linear programming tools (CPLEX) shows that Tolstoi’s solution indeed is optimum. But it is unclear how sure Tolstoi could have been about his claim that his solution is optimum. Geographical insight probably has helped him in growing convinced of the optimality of his solution. On the other hand, it can be checked that there exist feasible solutions that have none of the negative-cost cycles considered by Tolstoi in their residual graph, but that are yet not optimum.

2. Max-Flow Min-Cut

The Soviet rail system also roused the interest of the Americans, and again it inspired fundamental research in optimization.

In their basic paper *Maximal Flow through a Network* (published first as a RAND Report of November 19, 1954), Ford and Fulkerson [1954] mention that the maximum flow problem was formulated by T.E. Harris as follows:

Consider a rail network connecting two cities by way of a number of intermediate cities, where each link of the network has a number assigned to it representing its capacity. Assuming a steady state condition, find a maximal flow from one given city to the other.

In their 1962 book *Flows in Networks*, Ford and Fulkerson [1962] give a more precise reference to the origin of the problem:

It was posed to the authors in the spring of 1955 by T.E. Harris, who, in conjunction with General F.S. Ross (Ret.), had formulated a simplified model of railway traffic flow, and pinpointed this particular problem as the central one suggested by the model [11].


\[5\text{The maximum objective value of a feasible solution, whose residual graph does not contain any nonnegative-cost cycle of length 4, and not any of the seven longer nonnegative-length cycles considered by Tolstoi (of lengths 6 and 8), is equal to 397,226.}\]

\[6\text{There seems to be some discrepancy between the date of the RAND Report of Ford and Fulkerson (November 19, 1954) and the date mentioned in the quotation (spring of 1955).}\]

\[7\text{In their book, Ford and Fulkerson incorrectly date the Harris-Ross report October 24, 1956.}\]
written for the US Air Force. At our request, the Pentagon downgraded it to “unclassified” on May 21, 1999.

As is known (Billera and Lucas [1976]), the motivation for the maximum flow problem came from the Soviet railway system. In fact, the Harris-Ross report solves a relatively large-scale maximum flow problem coming from the railway network in the Western Soviet Union and Eastern Europe (‘satellite countries’). Unlike what Ford and Fulkerson say, the interest of Harris and Ross was not to find a maximum flow, but rather a minimum cut (‘interdiction’) of the Soviet railway system. We quote:

Air power is an effective means of interdicting an enemy’s rail system, and such usage is a logical and important mission for this Arm.

As in many military operations, however, the success of interdiction depends largely on how complete, accurate, and timely is the commander’s information, particularly concerning the effect of his interdiction-program efforts on the enemy’s capability to move men and supplies. This information should be available at the time the results are being achieved.

The present paper describes the fundamentals of a method intended to help the specialist who is engaged in estimating railway capabilities, so that he might more readily accomplish this purpose and thus assist the commander and his staff with greater efficiency than is possible at present.

First, much attention is given in the report to modeling a railway network: taking each railway junction as a vertex would give a too refined network (for their purposes). Therefore, Harris and Ross propose to take ‘railway divisions’ (organizational units based on geographical areas) as vertices, and to estimate the capacity of the connections between any two adjacent railway divisions. In 1996, Ted Harris remembered (Alexander [1996]):

We were studying rail transportation in consultation with a retired army general, Frank Ross, who had been chief of the Army’s Transportation Corps in Europe. We thought of modeling a rail system as a network. At first it didn’t make sense, because there’s no reason why the crossing point of two lines should be a special sort of node. But Ross realized that, in the region we were studying, the “divisions” (little administrative districts) should be the nodes. The link between two adjacent nodes represents the total transportation capacity between them. This made a reasonable and manageable model for our rail system.

The Harris-Ross report stresses that specialists remain needed to make up the model (which is always a good tactic to get a new method accepted):

The ability to estimate with relative accuracy the capacity of single railway lines is largely an art. Specialists in this field have no authoritative text (insofar as the authors are informed) to guide their efforts, and very few individuals have either the experience or talent for this type of work. The authors assume that this job will continue to be done by the specialist.

The authors next dispute the naive belief that a railway network is just a set of disjoint through lines, and that cutting these lines would imply cutting the network:

It is even more difficult and time-consuming to evaluate the capacity of a railway network comprising a multitude of rail lines which have widely varying characteristics. Practices among individuals engaged in this field vary considerably, but all consume a great deal of time. Most, if not all, specialists attack the problem by viewing the railway network as an aggregate of through lines.

The authors contend that the foregoing practice does not portray the full flexibility of a large network. In particular it tends to gloss over the fact that even if every one of a set of independent through lines is made inoperative, there may exist alternative routings which can still move the traffic.
This paper proposes a method that departs from present practices in that it views the network as an aggregate of railway operating divisions. All trackage capacities within the divisions are appraised, and these appraisals form the basis for estimating the capability of railway operating divisions to receive trains from and concurrently pass trains to each neighboring division in 24-hour periods.

Whereas experts are needed to set up the model, to solve it is routine (when having the 'work sheets'):

The foregoing appraisal (accomplished by the expert) is then used in the preparation of comparatively simple work sheets that will enable relatively inexperienced assistants to compute the results and thus help the expert to provide specific answers to the problems, based on many assumptions, which may be propounded to him.

For solving the problem, the authors suggested applying the ‘flooding technique’, a heuristic described in a RAND Report of August 5, 1955 by A.W. Boldyreff [1955a]. It amounts to pushing as much flow as possible greedily through the network. If at some vertex a ‘bottleneck’ arises (that is, more trains arrive than can be pushed further through the network), the excess trains are returned to the origin. The technique does not guarantee optimality, but Boldyreff speculates:

In dealing with the usual railway networks a single flooding, followed by removal of bottlenecks, should lead to a maximal flow.

Presenting his method at an ORSA meeting in June 1955, Boldyreff [1955b] claimed simplicity:

The mechanics of the solutions is formulated as a simple game which can be taught to a ten-year-old boy in a few minutes.

The well-known flow-augmenting path algorithm of Ford and Fulkerson [1955], that does guarantee optimality, was published in a RAND Report dated only later that year (December 29, 1955). As for the simplex method (suggested for the maximum flow problem by Ford and Fulkerson [1954]) Harris and Ross remarked:

The calculation would be cumbersome; and, even if it could be performed, sufficiently accurate data could not be obtained to justify such detail.

The Harris-Ross report applied the flooding technique to a network model of the Soviet and Eastern European railways. For the data it refers to several secret reports of the Central Intelligence Agency (C.I.A.) on sections of the Soviet and Eastern European railway networks. After the aggregation of railway divisions to vertices, the network has 44 vertices and 105 (undirected) edges.

The application of the flooding technique to the problem is displayed step by step in an appendix of the report, supported by several diagrams of the railway network. (Also work sheets are provided, to allow for future changes in capacities.) It yields a flow of value 163,000 tons from sources in the Soviet Union to destinations in Eastern European ‘satellite’ countries (Poland, Czechoslovakia, Austria, Eastern Germany), together with a cut with a capacity of, again, 163,000 tons. So the flow value and the cut capacity are equal, hence optimum.

In the report, the minimum cut is indicated as 'the bottleneck' (Figure 2). While Tolstoi and Harris-Ross had the same railway network as object, their objectives were dual.
Figure 2
From Harris and Ross [1955]: Schematic diagram of the railway network of the Western Soviet Union and Eastern European countries, with a maximum flow of value 163,000 tons from Russia to Eastern Europe, and a cut of capacity 163,000 tons indicated as “The bottleneck”.

Acknowledgements. I thank Sasha Karzanov for his efficient help in finding Tolstoi’s paper in the (former) Lenin Library in Moscow, Irina V. Karzanova for accurately providing me with an English translation of it, Kim H. Campbell and Joanne McLean at Air Force Pentagon for declassifying the Harris-Ross report, and Richard Bancroft and Gustave Shubert at RAND Corporation for their mediation in this.
References


[1930] A.N. Tolstoi, Metody nakhodjeniya naimen’shego summovogo kilometraza v prostranstve [Russian; Methods of finding the minimal total kilometrage in cargo-transportation planning in space], in: *Planirovanie Perevozok, Sbornik pervyi* [Russian; Transportation Planning, Volume I], Transpechat’ NKPS [TransPress of the National Commissariat of Transportation], Moscow, 1930, pp. 23–55.

[1939] A. Tolstoi, Metody ustraneniya neratsional’nykh perevozok pri planirovanii [Russian; Methods of removing irrational transportation in planning], *Sotsialisticheskii Transport* 9 (1939) 28–51 [also published as ‘pamphlet’: *Metody ustraneniya neratsional’nykh perevozok pri sostavlenii operativnykh planov* [Russian; Methods of Removing Irrational Transportation in the Construction of Operational Plans], Transzheldorizdat, Moscow, 1941].