Journal of Global Optimization 7: 381–405, 1995.

Bilevel Programming in Traffic Planning: Models, Methods and Challenge

ATHANASIOS MIGDALAS samig@math.liu.se
Division of Optimization, Department of Mathematics Linköping Institute of Technology, S-581 83 Linköping, SWEDEN

(Received: 30 December 1994; accepted: 15 May 1995)

Abstract. Well-founded traffic models recognize the individual network user's right to the decision as to when, where and how to travel. On the other hand, the decisions concerning management, control, design and improvement investments are made by the public sector in the interest of the society as a whole. Hence, transportation planning is a characteristic example of a hierarchical process, in which the public sector at one level makes decisions seeking to improve the performance of the network, while at another level the network users make choices with regard to route, travel mode, origin and destination of their travel. Our objective is to provide a review on the current state of research and development in bilevel programming problems that arize in this context, and attract the attention of the global optimization community to this problem class of imense practical importance.

Keywords: Bilevel programming, network problems, traffic planning.

1. Introduction

The movement of people and goods is as old as humanity itself. The need for transportation is an integrated part of any functioning society and stems from the interaction of social and economic activities. Society depends upon the mobility provided by transportation networks in order to make it possible for its members to participate in essential activities such as production, consumption, communication and recreation.

Approximately 70% of the citizens of the industrialized countries cluster today in cities and their metropolitan areas. Many of these metropolitan areas face serious congestion problems due to the increasing number of vehicles, which threaten to deteriorate the quality of life and increase air pollution. The environmental, economic, health and safety impacts are well known. It is estimated that only in the US, the traffic congestion in 1987 accounted for more than 2 billion vehicle hours of delay and that 2.2 billion gallons of excessive fuel consumption. The unavoidable dramatic increase in travel demand coupled with the diminishing construction of new transportation facilities will unavoidably worsen the traffic conditions unless some innovative congestion-relief methods can be developed and implemented in time. Thus, it is not surprising that tremendous energy and money are put in studies, research and development at regional, national and international level. Within the EU, for instance, the DRIVE, SOCRATES and lately the TELEMATICS programs are to large extent concerned with traffic questions.

The quantitative analysis of traffic phenomena yields network models that represent the spatial characteristics of the underlying infrastructure. These models are concerned with the prediction of the way in which vehicles use an existing or proposed infrastructure, or with the determination of the way in which such a utilization should be done. The presense of congestion is modeled with delay functions that are nonlinear, however, in most applications convex or monotone increasing. Thus, the resulting models are formulated as convex cost optimization network models and more general versions of these models as nonlinear complementarity or variational inequality problems with embedded network structures.

Well-founded models recognize the individual network user's right to the decision as to when, where and how to travel. On the other hand, the decisions concerning management, control, design and improvement investments are made by the public sector in the interest of the society as a whole. Hence, transportation planning is a characteristic example of a hierarchical process, in which the public sector at one level makes decisions seeking to improve the performance of the network, while at another level the network users make choices with regard to route, travel mode, origin and destination of their travel, etc. Thus, the responses of the network users can be predicted but not dictated. For example, society selects the links for capacity improvements but the users choose the routes they percieve to be best. Other examples include pricing of freigh transportation, traffic signal setting and origin-destination matrix estimation based on traffic counts, etc. Thus, the resulting models are formulated as bilevel programming models.

The bilevel programming problem describes a hierarchical system which is composed of two levels of decision makers. The higher level decision maker, known as *leader*, controls the decision variables $y \in Y$, while the lower level decision maker, known as *follower*, controls the decision variables $x \in X$. The interaction between the two levels is modelled in their respective *loss functions* $\varphi(x, y)$ and $f(x, y)$. The leader and the follower play a *Stackelberg duopoly game* [82], [6]. The idea of the game is as follows: The first player, the leader, chooses $y \in Y$ to minimize the loss function $\varphi(x, y)$, while the second player, the follower reacts to leader's decision by selecting a strategy $x \in X$ that minimizes his loss function $f(x, y)$, in full knowledge of the leader's decision. Thus, the follower's decision depends upon the leader's decision, i.e. $x = x(y)$, and the leader is in full knowledge of this. Consequently, we have the following definition:

Definition. If there exists a mapping $x : Y \to X$ such that for any fixed $y \in Y$,

$$f(x(y), y) \leq f(x, y), \quad \forall x \in X, \tag{1.1}$$

and if there exists $y^\star \in Y$ such that

$$\varphi(x(y^\star), y^\star) \leq \varphi(x(y), y), \quad \forall y \in Y, \tag{1.2}$$

then the pair $(x^\star, y^\star)$, where $x^\star = x(y^\star)$, is called a *Stackelberg equilibrium* with the first player as leader and the second player as follower.

Bilevel programming models are derived from the conditions (1.1)-(1.2) of the definition in a most natural way. Existence of Stackelberg equilibrium is guaranteed in the following case:

THEOREM 1.1 *If Y and X are compact sets in R^m and R^n respectively, and if φ and f are real-valued continuous functions on $X \times Y$, then a Stackelber equilibrium (with either player as leader) exists.[78]*

Consider the following bilevel program:

$$\textbf{(BP)} \min_{y \in Y} \quad \varphi(x(y), y) \tag{1.3}$$

$$\text{where } x(y) = \arg\min_{x \in X} f(x, y) \tag{1.4}$$

We assume here that the sets $X \subseteq R^n$ and $Y \subseteq R^m$ are convex and compact, that f and φ are continuous on $X \times Y$, and that (1.4) has a unique solution $x(y) \in X$ for all $y \in Y$. Thus, by Theorem 1.1, a Stackelberg equilibrium exists.

Problem (1.3)-(1.4) constitutes a generalization of most bilevel models in traffic planning, with the variables y of first level (1.3) corresponding to the decision variables of the society (network planner) and with the second level (1.4) being in the role of an oracle that supplies the leader with predictions concerning the reaction of the network users with respect to their mode, route, destination, etc. decisions.

1.1. Purpose

Stackelberg games and bilevel programming problems have been studied extensively in their general setting during the last decade [6], [5], [10], [11], [56], [4], [91], [92]. For a continuously updated bibliography on the subject see [95].

The purpose of the present paper is to provide information on the current state of the research in bilevel traffic planning problems, and attract the attention of the global optimization community to this problem class of immense practical importance. The resistance, due to inherent difficulty and size, of these problems to global optimization techniques constitutes a challenge to the community, especially as their bare existence requires their solution to global optimality. Indeed, nonproper solutions lead to both theoretical and real-world paradoxes, as for instance, in the case where increase in link capacities or construction of new links may deteriorate the performance of the network, i.e., it may yield increased travel times and bad network utilization [16], [54].

No attempt is made to cover all aspects of this large area of research and application development, we rather focus on a few basical classes of models and solution methods.

1.2. Outline

The paper is organized in the following way: First, the network equilibrium problem is presented and different formulations are given. Subsequently bilevel formulations of the network design, the signal setting and the origin/destination matrix adjustment problems are stated. In all three cases, the network equilibrium problem constitutes the second level. Finally, we overview the proposed solution techniques and classify them into exact methods, single-level approximations, bicriterion approximations, linear bilevel approximations, and local search methods. We conclude the paper with a few remarks in Section 4.

2. Models

In order to appreciate the practical importance and necessity of the bilevel programming models in traffic planning processes, one has to be conscious about the behavioral assumptions that govern the modelling process. Thus, although not in the realm of the (multiextremal) global optimization, instances of the network equilibrium problem are derived in section 2.1. This is the second level or followers' problem in all bilevel traffic planning problems of the subsequent sections. We restrict the discussion to single transportation mode.

2.1. The Network Equilibrium Problem

Traffic equilibria models are descriptive in the sense that their aim is to predict flow patterns and travel times which are the results of the network users' choices with regard to routes from their origins to their destinations. The input to the model is a complete description of the proposed or existing transportation system. The models are based on the behavioral assumption that "the journey times on all the routes used are equal, and less than those which would be experienced by a single vehicle on any unused route" [97]. This is *Wardrop's first condition*, also known as descriptive assignment or equal times journey principle. The traffic flows that satisfy this principle are usually referred to as *user-equilibrium* or *user-optimum* flows, since the routes chosen by the network users are those which are individually perceived to be the shortest under the prevailing conditions. The result from such individual decisions is a condition in which no user can reduce his/hers travel time by choosing unilaterally another route, i.e. it is an equilibrium condition in a noncooperative Nash-Cournot game where the players are associated with origin-destination pairs [50].

By contrast, *system optimum* flows satisfy *Wardrop's second condition* which states that "the average journey time is minimum". These flow patterns are characterized by the fact that all routes used between an origin and a destination have equal marginal travel times, that is, the total travel time in the network is minimized, and this is considered as the society's (the system's) understanding of optimal network utilization. However, the total travel time is generally *not* minimized by the user-optimal travel flows, and, moreover, observed flows in real life are closer to user than system optimum. The only situation in which the two flow patterns are equal is in the absence of congestion; this is an ideal case of course. For a thorough and up-to-date treatment of the network equilibrium problem see Patriksson[76].

2.1.1. Nonlinear Network Model

Traffic equilibrium problems are frequently divided into two modelling cases: *fixed demand* and *elastic demand*. In the fixed demand case, an origin-destination demand matrix $R = [r_k]$, with r_k being the travel demand between the kth origin-destination pair, is assumed given. By contrast, in the elastic demand case, the demand r_k is modelled as a function of the least travel cost between the end points of the kth origin-destination pair. Thus, the user has a number of travel choices available and he/she is economically motivated in his/her decisions of making or a not the trip.

Let $G = (A, N)$ be the underlying network with N the set of nodes and A the set of links, $K \subseteq N \times N$ the set of origin-destination pairs, P_k the set of simple paths (routes) between the end nodes of the kth origin-destination pair, c_{pk} and h_{pk} the travel time and flow respectively on the pth route in P_k.

According to Wardrp's equilibrium principle, if π_k denotes the shortest route travel time between the end nodes of the kth origin-destination pair, then

$$h_{pk} > 0 \implies c_{pk} = \pi_k, \ \forall p \in P_k, \tag{2.1}$$

$$h_{pk} = 0 \implies c_{pk} \geq \pi_k, \ \forall p \in P_k, \tag{2.2}$$

hold for all pairs in K. Thus, including flow feasibility constraints, the *user equilibrium conditions for fixed demand* can be stated as follows:

$$\begin{aligned} h_{pk}(c_{pk} - \pi_k) &= 0, \ \forall p \in P_k, & (2.3)\\ c_{pk} - \pi_k &\geq 0, \ \forall p \in P_k, & (2.4)\\ \sum_{p \in P_k} h_{pk} &= r_k, & (2.5)\\ h_{pk} &\geq 0, \ \forall p \in P_k & (2.6)\\ \pi_k &\geq 0, & (2.7) \end{aligned}$$

for all origin-destination pairs k.

For every link $a \in A$, let x_a denote the total link flow, and let $s_a(x_a)$ be the link travel cost encountered by a user travelling on link a with a total flow x_a. Define the link-route incidence matrix $\Delta = [\delta_{kap}]$, where δ_{kap} is 1 if the route p of the kth origin-destination pair uses link a, and 0 otherwise.

THEOREM 2.1 *Assume that the network $G = (N, A)$ is stronly connected with respect to the pairs in K, that the demand matrix R is nonnegative, and that the travel time function s_a is positive, strictly monotone increasing and continuously differentiable. Then, conditions (2.3)-(2.7) are the first-order optimality conditions of the convex optimization problem [24]*

$$\textbf{(FTAP)} \ \min \ \sum_{a \in A} \int_0^{x_a} s_a(t)dt, \tag{2.8}$$

subject to

$$\sum_{p \in P_k} h_{pk} = r_k, \ \forall k, \tag{2.9}$$

$$\sum_k \sum_{p \in P_k} \delta_{kap} h_{pk} = x_a, \ \forall a \in A \tag{2.10}$$

$$h_{pk} \geq 0, \ \forall p \in P_k, \forall k. \tag{2.11}$$

By contrast, the system optimum seeking problem can be stated as follows:

$$\textbf{(STAP)} \ \min \ \sum_{a \in A} s_a(x_a) x_a, \tag{2.12}$$

subject to (2.9)-(2.11).

Both problems are strictly convex in link flows, that is, the optimum link flow pattern is unique, however, they are only convex in route flows.

To extend the user equilibrium model formulation to the elastic demand case, let $r_k = g_k(\pi)$, where $\pi = [\pi_1, \ldots, \pi_k, \ldots]$, that is, the travel demand r_k between the end nodes of the kth origin-destination pair is a function of the vector of the cheapest route costs. Then Wardrop's user equilibrium principle for both route flows and demands are mathematically stated as follows:

$$h_{pk} > 0 \implies c_{pk} = \pi_k, \ \forall p \in P_k, \tag{2.13}$$

$$h_{pk} = 0 \implies c_{pk} \geq \pi_k, \ \forall p \in P_k, \tag{2.14}$$

$$r_k > 0 \implies r_k = g_k(\pi), \tag{2.15}$$

$$r_k = 0 \implies g_k(\pi) \leq 0, \tag{2.16}$$

for all pairs k.

Introducing the flow feasibility requirements, the above equations lead, under the additional requirement of nonnegative g_k on the nonnegative orthant, to the following conditions for an elastic demand user equilibrium:

$$h_{pk}(c_{pk} - \pi_k) = 0, \ \forall p \in P_k, \tag{2.17}$$

$$c_{pk} - \pi_k \geq 0, \ \forall p \in P_k, \tag{2.18}$$

$$\sum_{p \in P_k} h_{pk} = g_k(\pi), \tag{2.19}$$

$$h_{pk} \geq 0, \ \forall p \in P_k \tag{2.20}$$

$$\pi_k \geq 0, \tag{2.21}$$

for all origin-destination pairs k.

In the seminal work [12], Beckmann et al recognized in (2.17)-(2.21) an optimization problem. Indeed, assume that g_k has the additional property of being continuous and strictly decreasing. Then it is invertible, in which case $r_k = g_k(\pi_k) \iff \pi_k = g^{-1}(r_k)$, whenever $r_k > 0$.

THEOREM 2.2 *Assume that the network $G = (N, A)$ is stronly connected with respect to the pairs in K, and that the travel time function s_a is positive, monotone increasing and continuously differentiable. Then, conditions (2.17)-(2.21) are the first-order optimality conditions of the convex optimization problem [12]*

$$\textbf{(ETAP)} \ \min \ \sum_{a\in A}\int_0^{x_a} s_a(t)dt - \sum_k \int_0^{r_k} g_k^{-1}(t)dt, \tag{2.22}$$

subject to

$$\sum_{p\in P_k} h_{pk} = r_k, \ \forall k, \tag{2.23}$$

$$\sum_k \sum_{p\in P_k} \delta_{kap} h_{pk} = x_a, \ \forall a \in A \tag{2.24}$$

$$h_{pk} \geq 0, \ \forall p \in P_k, \forall k, \tag{2.25}$$

$$r_k \geq 0, \ \forall k. \tag{2.26}$$

2.1.2. Variational Inequality Formulation

While in the classical traffic equilibrium models the link travel functions s_a depend only on the link flow x_a, in the general case the link travel functions may also depend on the flow of neighbouring links, indeed on the entire link flow pattern $x = [x_1, \ldots, x_a, \ldots]$. The network equilibrium problem for such link functions was formulated as variational inequalities in the seminal papers [80], [22], [23] and as a nonlinear complementarity problem in [2].

Let $s(x) = [s_1(x), \ldots, s_a(x), \ldots]$ be the vector link travel function, $c(h) = [c_1(h), \ldots, c_k(h), \ldots]$ the vector of route costs, where $c_k(h)$ is the vector of route costs $[c_{kp}]$ for the kth origin- destination pair, i.e. $c_{kp} = \sum_{a\in A} \delta_{kap} s_a(x)$, and let $h = [h_1, \ldots, h_k, \ldots]$ be the vector of route flows, where h_k is the vector of route flows for the kth origin-destination pair.

The fixed demand equilibrium assignment is given by the following theorem.

THEOREM 2.3 *Assume that the network $G = (N, A)$ is stronly connected with respect to the pairs in K, that the demand matrix R is nonnegative, the route cost c is positive, continuously differentiable and monotone, and that the travel time s is positive, monotone and continuously differentiable. Then, the Wardrop conditions (2.3)-(2.7) are equivalent to the variational inequality problem of seeking an $h^* \in H$ such that*

$$c(h^*)^T(h - h^*) \geq 0, \ \forall h \in H, \tag{2.27}$$

where H is the polyhedron defined by (2.9)-(2.11). Moreover, the Wardrop conditions (2.3)-(2.7) are also equivalent to the variational inequality problem of seeking an $x^* \in X$ such that

$$s(x^*)^T(x - x^*) \geq 0, \ \forall x \in X, \tag{2.28}$$

where X is the polyhedron of feasible link flows implied by (2.9)-(2.11).

Next we consider the elastic demand formulation of the problem. Let the demand function be $g_k(\pi)$, where π is the vector of shortest route travel times for all origin-destination pairs of the network, and consider the vector demand function $g(\pi) = [g_1(\pi), \ldots, g_k(\pi), \ldots]$.

THEOREM 2.4 *Assume that the network $G = (N, A)$ is strongly connected with respect to the pairs in K, that the route cost c is positive, continuously differentiable and monotone, and that the travel time s is positive, monotone and continuously differentiable. If the demand function g is bounded from above, continuously differentiable, nonnegative, strictly monotone, and invertible, then the Wardrop conditions (2.17)-(2.21) are equivalent to the variational inequality problem of seeking an $(h^*, r^*) \in H_r$ such that*

$$c(h^*)^T(h - h^*) - g^{-1}(r^*)^T(r - r^*) \geq 0, \; \forall (h, r) \in H_r, \tag{2.29}$$

where H_r is the polyhedron defined by (2.23)-(2.26) and r is the demand vector $[r_k]$.

When the link travel functions are strictly monotone and the demand function is strictly decreasing, the network equilibrium model has unique link flows, demands and origin to destination costs. This is an important result, as it allows comparisons between different potential configurations of the network.

2.2. The Network Design Problem

The network design problem is concerned with the improvement of a transportation system through modification of link capacities, including addition (and sometimes deletion) of links. Since it is the public sector which is concerned with the management of the transportation system, the objective is to minimize *total system costs* consisting of system travel cost in the sense of section 2.1, and investment. On the other hand, the system cannot prescribe the users' behavior and therefore the total travel time is computed by evaluating the objective function (2.12) for user-equilibrium flows which have to be predicted by solving the network equilibrium problem. Thus, only the investment costs are controlled and allocated in an optimal way from the system's perspective. However, although unable to prescribe routes to the users, the system influences their choices by selecting subsets of links for improvements and making them more attractive than others.

The bilevel formulation of the network design problem is due to LeBlanc, who in the seminal paper [60] presented a discrete case of the problem, and subsequently the continuous case in [3]. Only the continuous case is considered here.

Let y_a be the modification (raising or lowering) made to the capacity of link $a \in A$ and denote y the corresponding vector. The link travel time s_a is, of course, influenced by such capacity modifications. Therefore, $s_a = s_a(x_a, y_a)$. Let $\psi_a(y_a)$ be the investment on link $a \in A$, and l_a and u_a the lower and upper levels of allowed capacity modification.

The network design problem can then be stated as follows:

$$\textbf{(NDP)} \min_{y} \quad \sum_{a \in A} s_a(x_a, y_a)x_a + \sum_{a \in A} \psi_a(y_a) \tag{2.30}$$

subject to

$$l_a \le y_a \le u_a, \ \forall a \in A \tag{2.31}$$

where x_a are the user-equilibrium link volumes predicted by solving the problem (c.f. section 2.1)

$$\min_{x} \quad \sum_{a \in A} \int_0^{x_a} s_a(t, y_a)dt \tag{2.32}$$

subject to (2.9)-(2.11).

THEOREM 2.5 *For fixed y_a, assume that the function $s_a(x_a, y_a)$ is continuously differentiable, positive and strictly monotone increasing for all $x_a \ge 0$. Then, the second level problem (2.32) has a unique optimal solution [2].*

A typical variant of the model is obtained by replacing the first level (2.30)-(2.31) with its budget-constrained counterpart, i.e.,

$$\textbf{(BNDP)} \min_{y} \quad \sum_{a \in A} s_a(x_a, y_a)x_a \tag{2.33}$$

subject to

$$\sum_{a \in A} \psi_a(y_a) \quad \le \quad b \tag{2.34}$$

$$l_a \le y_a \le u_a, \ \forall a \in A, \tag{2.35}$$

where b is the available budget. Other variants may be obtained by replacing the second level problem.

The functional form of ψ_a varies with the situation that is modelled. Usually linearity (e.g. [49]) or convexity (e.g. [64]) is assumed. A number of specialized functional forms are studied in [65]. However, the convexity of ψ_a is questionable from the practical point of view, since, as observed in [3], convex investment functions result in minor modifications in the capacity of many arcs. Moreover, as discussed in [83], [84], economy-of-scale or s-shaped costs (i.e. d.c.-functions) are present in almost all real life applications.

The *system optimum network design problem* is a single-level optimization problem in which the users' behavior is ignored:

$$\textbf{(SNDP)} \min_{y} \quad \sum_{a \in A} s_a(x_a, y_a)x_a + \sum_{a \in A} \psi_a(y_a) \tag{2.36}$$

subject to

$$l_a \le y_a \le u_a, \ \forall a \in A \tag{2.37}$$

and (2.9)-(2.11).

The functions $s_a(x_a, y_a)$ is usually assumed to have the form $s_a(\frac{x_a}{y_a})$, conforming to the so-called BPR[1] function type, and to be positive, increasing function of the ratio $\frac{x_a}{y_a}$ and continuously differentiable. By conversion, $s_a(\frac{x_a}{y_a}) = 0 \Longleftrightarrow x_a = y_a = 0$.

THEOREM 2.6 *Under the above assumptions, function $x_a s_a(x_a, y_a)$ is convex. Moreover, if $\psi_a(y_a)$ is linear or convex, problem (2.36)-(2.37) is a convex programming problem.*

2.3. The Signal Setting Problem

In contrast to the network design, the signal control is used as a tool to increase the performance of the transportation system *without* changes in the infrastructure. Its purpose is to promote safety, efficiency and convenience in the mobility of people and goods through better utilization of the existing infrastructure.

The control adopted by the system influences the traffic pattern, that is, the signal control influences the values perceived by the users as travel costs and therefore induces new flow equilibria in the transportation networks. Such an effect can be achieved by changing link travel functions and delays at junctions.

Following Gartner et al [47], [52], [19] formulate the signal setting problem (**SSP**) as a bilevel program, where the first level concerns the decision of the public network manager who establishes the control system, and the second level concerns the behavior of the network users in their choice of routes. This formulation respects the fact that the manager's decisions are based on the public interest, while the individual user is concerned only with his/her own travel costs. See also [64], [32], [33].

Since the design of the control can be interpreted as a way of manipulating the link capacities of the network, the mathematical programming model is equivalent to that for the network design problem (2.30)-(2.32). However, since signal control is less capital intensive, investment costs do not usually enter in the objective or the constraints. Interpreting $y = [y_1, \ldots, y_a \ldots]$ as the network control system parameters, the manager is interesting in determining a *mutually consistent* user-equilibrium flow and control strategy.

2.4. The Origin/Destination Matrix Adjustment Problem

In most transportation planning applications, the input data which is mostly difficult and expensive to obtain is the origin-destination (O-D) demand matrix. This is so because the demand data is not directly observeable, on the contrary, it requires extensive and expensive surveys which involve home and road based interviews. On the other hand, link volumes are easily obtainable within reasonable precision by simply counting the traffic at certain count-posts, either manually or automatically.

Consequently, the problem of estimating or adjusting an O-D matrix from observed traffic flows has attracted considerable attention [72], [74], [94], [35], [36]. These studies propose a large variety of models, which can be classified according to the way that the observed data are used in the modelling and the way in which the O-D matrix is distributed over the paths of the network.

As discussed by Fisk [35], [36], the model with flow dependent link travel times is a bilevel programming problem, **ODP** below. This is a consequence of the equilibrium assumption. The observed link flows α do not, in general, satisfy exactly these equilibrium conditions, and therefore one is interested in finding an equilibrium flow pattern x that is close to the observed flows. However, the problem differs from those in section 2.2 and 2.3, since the leader is *not* trying to influence the network users. The second level program has here the bare role of a pure oracle.

The formulation of the problem is as follows:

$$\textbf{(ODP)} \min_{r} \quad \frac{1}{2}\sum_{a\in\hat{A}}(x_a(r)-\alpha_a)^2 \tag{2.38}$$

$$\text{s.t} \quad 0 \le r_k \le u_k, \forall k \in K \tag{2.39}$$

$$\text{where } x(r) \text{ solves (c.f. } \textbf{ETAP}) \quad \min_{x}\sum_{a\in A}\int_0^{x_a} s_a(t)dt \tag{2.40}$$

$$\text{s.t.} \sum_{p\in P_k} h_{pk} = r_k, \forall k \in K \tag{2.41}$$

$$h_{pk} \ge 0, \forall p \in P_k, \forall k \in K \tag{2.42}$$

$$x_a = \sum_{k\in K}\sum_{p\in P_k} \delta_{kap} h_{pk}, \forall a \in A \tag{2.43}$$

Here, $\hat{A}$ denotes the subset of the arcs for which flow counts α_a are available.

THEOREM 2.7 *Let $s_a(x_a)$ denote the link cost for arc a and assume it to be continuously differentiable and strictly monotone increasing. Then, for fixed r, the second level program (2.40)-(2.43) has a strictly convex objective function and, if feasible, a unique optimal solution $x(r)$.*

Since the O/D matrix adjustment problem as formulated above admits an infinite number of optimal solutions [81], there is a second version of the **ODP** problem in which a regularization, in the sense of [88], [68], term is added to the objective of the first level problem, i.e.

$$\textbf{(ODP')} \min_{r} \quad \frac{1}{2}\sum_{a\in\hat{A}}(x_a(r)-\alpha_a)^2 + \frac{1}{2}\sum_{k\in K}(r_k-\beta_k)^2 \tag{2.44}$$

Here β is a target O-D matrix.

3. Solution Methods

Real-world cases yield traffic planning problems of impressive sizes. For example, the network of the (part of) city of Barcelona, Spain, consists of 1020 nodes, 2522 links and 7922 origin-destination pairs. The city of Linköping, Sweden, which is far smaller than Barcelona but of large spatiality has a network consisting of 300 nodes, 400 links and 7000 origin-destination pairs. Finally, the network of Winnepeg, Canada, has 1052 nodes, 2836 links and 4344 origin-destination pairs. The sizes of the resulting problems are left as an exercise to the reader.

Despite the enormous sizes, the convex programming problems of section 2.1, are efficiently solved through maximal utilization of the underlying problem structures, i.e. spatiality, sparsity and network (see e.g. Patriksson [76] for an extensive survey of methods). In particular, utilization of a column generation scheme, called Simplicial Decomposition [51], [57], [59], is able to solve these problems within a few minutes on sequential machines, while the Distributed Simplicial Decomposition [26] solves the problems within a few seconds on parallel machines. Methods based on the node-link representation of the problems are also quite efficient (see e.g. [62], [58], [28]).

On the other hand, the bilevel programming problems in sections 2.2 to 2.4 are in another complexity class. Indeed, it has been shown in [53], [13], [9], [48] that even the linear case of the bilevel programming problem is NP-hard. Moreover, the studies [8], [10], [4], [91], [92] reveal the fact that when the first order optimality conditions of the second level problem are both necessary and sufficient, problem **BP** can be solved as a global optimization problem.

Subsequently we overview the methods proposed for the solution of the problems in sections 2.2 to 2.4, as well as a few that potentially could be used for this purpose. The discussion is based on the model **BP**, however, whenever necessary direct reference to problems in sections 2.2 to 2.4 is made.

3.1. Exact Methods

Very few exact methods have been proposed for the nonlinear bilevel programming problem in general and the traffic planning cases in particular. In their current state of development, these methods are unable to attack real-world cases of the size exemplified above with the Linköping, Barcelona and Winnipeg networks. The inherent difficulty of the bilevel problem is one reason, however, a second important reason is that the underlying spatiality, sparsity and network structure of the problems are not incorporated in the derivation of the methods and consequently not exploited during the solution process, that is, the methods are of too general character.

3.1.1. Replacement of the Second Level with Karush-Kuhn-Tucker Conditions

One of the first attempts to solve **BP** to global optimality was made in [10]. The second level problem is replaced by its first order optimality (Karush-Kuhn-Tucker, **KKT**) conditions. Under suitable regularity assumptions, **KKT** are both necessary and sufficient.Thus, the obtained single level program was shown equivalent to **BP**. Under the additional assumption of function separability, a branch-and-bound algorithm was proposed to find the global optimum. For the strictly convex case, [8] presented a more efficient branch-and- bound. In [4], it is demonstrated that the complicating complementarity constraint in **KKT** can be replaced by an equivalent system of convex quadratic constraints, and that **BP** can be reduced to the equivalent problem of minimizing a concave objective function subject to a convex feasible region. A branch-and-bound algorithm is developed.

3.1.2. Reformulation to Single Level d.c. Program

In [70], following the technique of [91], [92], the optimal objective function value of **BP**, $v(y) = \min_{x \in X(y)} f(x,y)$, is used to reduce **BP** to an equivalent single level problem

$$\min_{y \in Y, x \in X(y)} \quad \varphi(x,y) \tag{3.1}$$

$$\text{subject to } v(y) \geq f(x,y). \tag{3.2}$$

A branch-and-bound algorithm, based on Theorem 3.1 below, is proposed for this problem.

THEOREM 3.1 *The value function v is convex in y. Consequently the constraint (3.2) is a* d.c. constraint.

3.1.3. Variational Inequalities in the Second Level

Reformulation of **BP** to an equivalent single level program is succeeded in [64] by replacing the second level program with its first order variational inequality conditions for optimality (c.f. section 2.1). Thus, **BP** is replaced by

$$\min_{y \in Y, x \in X} \quad \varphi(x,y) \tag{3.3}$$

$$\text{subject to } \nabla_x f(x,y)^T (x - x^q) \leq 0, \ \forall x^q \in X. \tag{3.4}$$

In the network design case, it is sufficient to consider in (3.4) only extreme points of the flow polytope, that is, only a finite number of constraints corresponding to x^q with $q \in \mathcal{E}(X)$. The proposed exact algorithm starts with a restricted number of (3.4)-constraints, and successively generates new as needed. The extreme point generator is the linearized (in the Taylor sense) approximation of the second level program, i.e.

$$\min_{x \in X} \quad \nabla_x f(\hat{x}, \hat{y})^T x, \tag{3.5}$$

where $(\hat{x}, \hat{y})$ is the solution to the current restriction of problem (3.3)-(3.4). Notice the similarity of the approach to the well-known Benders scheme.

3.2. Heuristics Based on Single Level Optimization

In contrast to the poor development of exact methods, the set of heuristics for the **BP** in general and the bilevel traffic planning problems in particular is very rich. Many of these heuristics have been developed to attack **BP** indirectly, i.e. by solving (perhaps to global optimality) other related single level problems. The replacement of **BP** is not always explicit. A number of such heuristics are gathered and analysed in [65].

3.2.1. System Optimum

The most straightforward approach of obtaining a single level optimization problem in the case of traffic planning is to ignore user behavior. For instance, in the case of network design, problem **SNDP** instead of **NDP** is solved. Of course, such an approach includes the danger for the occurence of traffic paradoxes[16], [54] and practically ignores the original problem totally. Therefore the development of such traffic planning approaches are dated to the era of 70's or to the noncongested case (see e.g. [18], [61], [83], [73], [84], [17], [74]).

3.2.2. Mixed Approach

In an attempt to circumvent the drawbacks of system optimum network design, [77] considers a mixed modelling approach. A single level approximation to **NDP** is introduced by adding the investment costs to the objective function of the second level problem. That is, the following problem instead of **NDP** is solved,

$$\min_{x,y} \quad \sum_{a \in A} \int_0^{x_a} s_a(t, y_a)dt + \sum_{a \in A} \psi_a(y_a) \tag{3.6}$$

subject to (2.31) and (2.9)-(2.11).

Of course, the optimum solution is neither system optimum nor does it obey the assumptions on the user behavior.

3.2.3. Block Coordinate Descent

The most popular heuristic algorithm for bilevel traffic planning problems resembles the well-known coordinate descent method. For fixed $y^q \in Y$, the second level problem in **BP** can be solved for the optimal solution $x^q = x(y^q) \in X$. Then, for this $x(y^q)$, the first level problem can be solved for a new $y^{q+1} = y(x(y^q))$. Repeating the entire process for $q = 0, 1, \ldots$, it will converge, under suitable assumptions, to a point (x^*, y^*), which, however, needs not be the global optimum in **BP**, as intuitively cautioned in [83], [84]. This approach was suggested by [87] in the context of **SSP**, who also gave numerical evidence for the non-global optimality of (x^*, y^*). Subsequently the heuristic was analyzed in [64], [65], where it is shown to actually solve a certain single level convex optimization problem (c.f. **FTAP** and **ETAP**). In [32], a numerical example is given where the algorithm converges to a Nash and not a Stackelberg equilibrium. See also [43], [49], [34], [52], [19], [101].

3.3. Bicriterion Approach

Following [7], [93] who proposed a bicriterion approach for the linear bilevel problem, [63] proposed a similar approach for the case of the traffic network design problem **NDP**. They replace **BP** with the parametrized problem

$$\min_{x \in X, y \in Y} \quad \gamma\varphi(x, y) + (1 - \gamma) f(x, y), \tag{3.7}$$

where $\gamma \in [0, 1]$. Moreover, in order to fully adapt the approach in [7], LeBlanc and Boyce [63] developed a piecewise linear approximation to **NDP**. Thus, the problem they were solving was a linear bicriterion problem. However, the approach was demonstrated inadequate in [13], [98], [66] with the means of counter examples. Deriving conditions under which an efficient point (Pareto optimum) to (3.7) also is a Stackelberg equilibrium in **BP** is of importance here. The following result is from [69].

THEOREM 3.2 *Consider the bicriteria problem*

$$\min_{x \in X,\, y \in Y} \{\varphi(x, y), f(x, y)\}, \tag{3.8}$$

and assume that $\forall y \in Y$, $x(y)$ *is a unique solution of (1.4), and that* $f(x, y) \leq f(\hat{x}, \hat{y}) \Rightarrow \varphi(x, y) < \varphi(\hat{x}, \hat{y}),\quad x, \hat{x} \in X, y, \hat{y} \in Y$. *Then, if* $(x(y^\star), y^\star)$ *is a solution to (1.3)-(1.4), then it is also an efficient solution to the vector problem (3.8). Conversely, if* $(x^\star, y^\star)$ *is an efficient point of (3.8) and* $x^\star$ *is a solution to (1.4) for fixed* $y = y^\star$, *then* $y^\star$ *is a solution to (1.3).*

Unfortunately, the typical functions involved i traffic planning do not satisfy the theorem's conditions.

3.4. Linear Bilevel Model Approximation

Under the assumption that bilevel linear programs are "easier" than their nonlinear convex counterparts, linear bilevel approximations to **BP** have been proposed. LeBlanc and Boyce [63], assuming linear investment functions in **NDP**, linearize (2.30)-(2.32) to

$$\min_{y} \quad \sum_{a\in A}\sum_{m\in M_a} c_{am}x_{am} + \sum_{a\in A}\psi_a y_a \tag{3.9}$$

$$\text{subject to} \quad y_a \geq 0, \forall a \in A \tag{3.10}$$

$$\text{where} \quad x = \arg\min_{x} \sum_{a\in A}\sum_{m\in M_a} \sigma_{am}x_{am} \tag{3.11}$$

$$\text{subject to} \quad \sum_{a\in S(i)} x_{ak} - \sum_{a\in T(i)} x_{ak} = r_{ik},\ \forall i \in N, \forall k \in K \tag{3.12}$$

$$\sum_{m\in M_a} x_{am} = \sum_{k\in K} x_{ak}, \forall a \in A \tag{3.13}$$

$$0 \leq x_{am} \leq u_{am} + \rho_{am}y_a,\ \forall m \in M_a, \forall a \in A \tag{3.14}$$

$$x_{am} \geq 0,\ \forall a \in A,\ \forall k \in K. \tag{3.15}$$

Here M_a is the set indexing the segments (c.f. c_{am}) in the piecewise linear approximation of total travel cost of link a, $\rho_{am}y_a$ are the capacity units received by the mth segment of link a, ρ_{am} is the prespecified proportion of added capacity to the mth segment of link a, x_{am} is the total flow volume of the mth segment of link a, u_{am} is the capacity of the mth segment of link a, c_{am} and σ_{am} are the coefficients of the mth piece of link a for the system and the cumulative user travel costs respectively, x_{ak} is the flow volume of the kth origin-destination pair on link a, x_{am} is the flow on the mth piece of link a, $S(i)$ is the set of link starting at node i, $T(i)$ is the set of links terminating at node i, and r_{ik} is r_k if i is the origin node of the kth pair, $-r_k$ if i is the destination node of the kth pair, and 0 otherwise. Equations (3.12) are the node-arc flow balancing equations that correspond to the arc-path formulation in **FTAP**.

The solution approach to (3.9)-(3.15) proposed in [63] consists of replacing the bilevel linear program by a parametrized single-level linear program as in (3.7), and it was based on the false conjecture that bilevel and bicriterion optimization are equivalent. Subsequently, Ben-Ayed et al [14] extended the reformulation (3.9)-(3.15) to include nonlinear, convex and concave, link investment functions. Application of the piecewise linear approximation approach to the Tunisian inter-regional highway network is presented in [15].

3.5. Bilevel Local Search

The algorithms discussed here constitute direct, but local, approaches to the solution of **BP**, that is, in the spirit of traditional nonlinear programming, they compute a stationary point, hopefully a local optimum, to **BP**. From the point of view of traffic planning, we can say that they respect the user equilibrium principle. Unfortunately, checking local optimality in bilevel programming, even in the all-linear case, is NP-hard [96]. Thus, there are no a priori guarantees on the quality of the produced point.

We classify the algorithms into three classes; direct search, penalty function and descent search methods. See also [55], [95].

3.5.1. Direct Search

The basic idea here is to accept the presence of the implicit functional $x = x(y)$ in leader's problem (1.3) and base the search for a stationary point y^s to (1.3) only on function evaluations. Each evaluation of $\varphi(x(y), y)$ would, of course, require the exact solution of the follower's problem (1.4). This means, in particular, that for the traffic planning problems of sections 2.2 to 2.4, a network equilibrium problem of some version in section 2.1 must be solved to optimality each time the system's objective function is evaluated. Although each instance of the network equilibrium problem can be solved quite efficiently, as mentioned in section 3, the number of repeated solutions required by a direct search method (e.g. the Hook-Jeeves method [3]) soon becomes computationally prohibitive with increasing problem sizes. Moreover, in the presence of nondifferentiability, both the theoretical and practical reputation of direct search methods is not the best (see e.g. [71]). Reduction in the number of network equilibrium problems that need to be solved is achieved in [86].

Friesz et al [46] try to overcome the tendency of the direct search to be trapped at narrow valleys and stationary points by introducing randomization of the trial points and embedding the method into an annealing scheme [1]. The Achilles heel of the simulated annealing approach is the identification of a good annealing schedule and the enormous amount of follower problems that have to be solved, one for each trial point. Friesz et al [46] partially circumvent this obstacle in the case of the network design problem (3.3)-(3.4) by storing path information generated during the search of first few hundreds or thousands of trial points and use them subsequently in reducing the effort needed for the solution of each of the follower problems.

3.5.2. Penalty Function Methods

For the general problem see the bibliography of [95]. In traffic planning, penalty methods have been proposed by e.g. Fisk [32], [33], particularly for the **SSP** problem. Her approach is based on the variational inequality (c.f. section 2.1) formulation of the follower's problem (c.f. (3.3)-(3.4)). Thus, **BP** takes the form

$$\min_{x \in X, y \in Y} \quad \varphi(x, y) \tag{3.16}$$

$$\text{subject to} \quad s(x,y)^T(z-x) \geq 0, \ \forall z \in X \tag{3.17}$$

Problem (3.16)-(3.17) can be simplified. For this, the following two theorems are required.

THEOREM 3.3 *Assume that $s(x,y)$ is monotone in x for any fixed $y \in Y$, then (3.17) is equivalent to the following max-min problem [38]*

$$\max_{x \in X} \ \min_{z \in Y} \ s(x,y)^T(z-x) \tag{3.18}$$

THEOREM 3.4 *Let*

$$w(x,y) \ = \ \min_{z \in X} s(x,y)^T(z-x), \tag{3.19}$$

then x^* is an optimal solution to (3.18) if, and only if, $w(x^*,y) = 0$. Moreover, if x is not a solution, then $w(x,y) < 0$ [104]

Function (3.19) is a so-called *gap function.* It helps in restating (3.16)-(3.17) in the simplified form below:

$$\min_{x \in X, y \in Y} \quad \varphi(x,y) \tag{3.20}$$

$$\text{subject to} \quad w(x,y) = 0 \tag{3.21}$$

Fisk's approach is based on penalizing constraints (3.21). A second reformulation of problem (3.16)-(3.18), equivalent to (3.20)-(3.21), is possible, if a gap function $\omega(x,y)$, also attributed to [104], is utilized in place of $w(x,y)$. The gap function $\omega(x,y)$ is obtained by interchanging the order of *min* and *max* in Theorem 3.3. Then Theorem 3.4 is still valid, however, $\omega(x,y) > 0$ whenever x is not a solution to (3.18).

3.5.3. Descent Search

The methods in this class are designed to compute stationary points (and hopefully local optima) to **BP** by searching along directions which are descent with respect to the leader's objective (1.3).

The derivative information on the implicit functional $x = x(y)$, needed for the computation of the descent direction, is obtained by various methods from the follower's problem. In particular, the work of Fiacco [30], [31] and Tobin [89], [90] on sensitivity analysis of nonlinear programs and variational inequalities respectively, constitutes the basis for derivative calculations. In the seminal paper [56], Kolstad and Lasdon put the basis for the development of descent search algorithms.

However, there are obstacles when trying to apply these results directly to traffic planning problems. In particular, as mentioned in section 2.1, the optimum route flows in the equilibrium network problem are not unique, even if the optimal link flows are. Thus, they do not meet the second-order sufficient condition for a local isolated point. Therefore, the results of Fiacco and Tobin are not applicable without modifications. See also [44], [45], [21], [85], [29].

The generic scheme of bilevel descent direction algorithms is given below. It involves two basic components: (i) finding a direction of descent, and (ii) computing an optimal step length. Since each evaluation of the leader's objective function requires the solution of a network equilibrium problem, step length calculations constitute the more difficult component of a descent algorithm.

! Step 0 : Initialize
Let $y^0 \in Y$ *! Feasible point*
$\varphi^0 = \varphi(x(y^0), y^0)$ *! Objective function value*
$q = 0$ *! Iteration counter*
repeat
    ! Step 1 : Determine a direction of descent
    $d_y^0 \in \mathcal{D}(y^0)$ *! $\mathcal{D}$ denotes the cone of feasible directions*
    ! Step 2 : Determine the optimal step length
    $\varphi(y^q + p_q d_y^q) = \min_{y:==y+pd_y^q \in Y} \varphi(x(y), y)$
    ! Step 3 : New iterate
    $y^{q+1} = y^q + p_q d_y^q$
    $\varphi^{q+1} = \varphi(x(y^{q+1}), y^{q+1})$
    $q = q + 1$
until(y^q *stationary*) *! Termination*

Algorithms in this category have been proposed for the network design problem by [45] for the case in which the follower's problem is formulated as variational inequalities (c.f. section 2.1 and (3.16)-(3.17)) and by [85] for the case in which follower's problem is formulated as an optimization problem. Sensitivity analysis based on **KKT** optimality conditions of the second level program is utilized to derive an equation system for the calculation of the derivatives of the implicit functional $x = x(y)$. Descent algorithms based on quasi-Newton methods, and particularly BFGS, are developed. In [85] the Armijo step length rule is used. In an attempt to reduce excessive step length computations, [45] use predetermined step lengths. However, convergence of such heuristics cannot be guaranteed.

For the signal setting problem **SSP**, descent algorithms based on sensitivity analysis of the second level problem are proposed in [99], [100], [103].

Chen and Florian [21] consider a reformulation of **BP** into a single level optimization problem of the form (3.1)-(3.2), however, with equality in (3.2). In particular, for the **ODP'**, they derive the following important theorem (see also [40], [39]).

THEOREM 3.5 *Under the standard assumptions on $s_a(x)$, $v(r)$ defined as the optimal objective value of the second level problem (2.40)-(2.43) is Gâteaux differentiable with $\nabla v(r) = \lambda(r)$, where $\lambda(r)$ is the optimal Lagrangean multiplier of the* *second level problem for a given r. Moreover, $v(r)$ is a monotone and continuous function of r.*

This result is used in [40] to develop a descent algorithm for **ODP'** by relaxing the equality form of (3.2) and attacking the resulting Augmented Lagrangean problem through linearization. The algorithm is designed to work with link rather than route flows. An Armijo step length rule is used for the determination of the step length. Under certain assumptions, convergence to a stationary point of the **ODP'** problem is demonstrated. The same authors propose in [39] another descent algorithm based on linearization of an Augmented Lagrangean problem, however, the complementarity equations of the **KKT** optimality conditions for the second level problem are relaxed instead of the equality (3.2). Subsequently, [41] realized that $\nabla v(r)_k = \sum_{a \in A} \gamma_{ak} s_a(x_a), \forall k \in K$, where γ_{ak} are the arc flow proportions of the demands r_k, defined by $\sum_{p \in P_k} \delta_{kap} \gamma_{kp}$ with γ_{kp} being the proportion of demand r_k assigned to path p, i.e. $h_{pk} = \gamma_{pk} r_k$. Based on this fact, they propose an *almost* descent algorithm which in each iteration alternates between direction generation for the first level of **ODP'** with γ fixed and link flow (i.e. γ) generation by solving a corresponding (3.1)-(3.2) for fixed demands r_k and equality in (3.2). They show that if the algorithm terminates, then the final point is a stationary point of the original problem **ODP'**. Their analysis hold also for the algorithms of [81], [101]. See also [102].

For the **NDP** problem, Davis [27] developed reduced gradient and sequential quadratic programming algorithms based on derivative information which is obtained from the second level program by replacing the deterministic modelling of section 2.1 with the so-called *Stochastic User Equilibrium Assignment.* In this setting, it is accepted that an individual's perceived travel costs are subject to random error. Consequently, some trips are assigned to routes having greater than minimal travel costs. However, the probability of a trip being assigned to a route is higher for lower cost routes. If $\gamma_{ak}(s)$ denotes the function which gives the probability of a trip between the kth origin-destination pair being using the link a and if $\gamma_{kp}(s)$ denotes the probability of choosing the pth route then $\gamma_{ak}(s) = \sum_{p \in P_k} \delta_{kap} \gamma_{kp}(s)$. Daganzo [25] has shown that $x_a = \sum_{k \in K} r_k \gamma_{ak}(s), \forall a \in A$, where the probabilities γ_{ak} are computed through the *probit model* (see e.g. [79]). Davis uses these equations in formulating a single level version of the network design problem, which he then solves with the aforementioned methods.

4. Some Remarks

Much progress has been made in understanding and modelling of hierarchical decision making and similar phenomena that arize during traffic planning processes. However, satisfactory solution methods for the resulting bilevel problems have not yet been developed. Although some progress has been made in this direction since the time for a similar conclusion by Friesz[42], currently, what are considered as the most promising methods are heuristics which do not even guarantee the local optimality of the final results. Moreover, the computational efford required by most of these methods is excessive, and therefore there is a considerable lack of computational results for networks of the sizes mentioned in section 3. Thus, the derivation of efficient algorithms for the solution of bilevel traffic planning problems is both a theoretical and practical challenge.

Acknowledgements

This work has been supported by the Swedish Communication Research Board (KFB).

Notes

1. Bureau of Public Roads

References

1. E. Aarts and J. Korst (1990) *Simulated Annealing and Boltzmann Machines - A Stochastic Approach to Combinatorial Optimization and Neural Computing*, John Wiley & Sons, Chichester
2. H. Aashtiani and T. Magnanti (1981) Equilibria on a congested transportation network, *SIAM Journal on Algebraic and Discrete Methods*, vol. 2, pp. 213-226
3. M.S. Abdulaal and L.J. LeBlanc (1979) Continuous Equilibrium Network Design Models, *Transportation Research*, vol. 13B, pp. 19-32
4. F.A. Al-Khayyal, R. Horst and P.M. Pardalos (1992) Global Optimization of Concave Functions subject to Quadratic Constraints: An Application in Nonlinear Bilevel Programming, In :: [5], pp. 125 - 147
5. G. Anandalingam and T.L. Friesz (1992) (Ed.s) *Hierarchical Optimization*, Annals of Operations Research, vol. 34, J.C. Baltzer AG, Basel.
6. J.P. Aubin (1979) *Mathematical Methods of Game and Economic Theory*, North-Holland, Amsterdam.
7. J.F. Bard (1983) An Efficient Point Algorithm for a Linear Two-Stage Optimization Problem, *Operations Research*, vol. 31, pp. 670-684
8. J.F. Bard (1988) Convex Two-Level Optimization, *Mathematical Programming*, vol. 40, pp. 15-27
9. J.F. Bard (1991) Some Properties of the Bilevel Programming Problem, *Journal of Optimization Theory and Applications*, vol. 68, pp. 371-378
10. J.F. Bard and J.E. Falk (1982) An Explicit Solution to the Multilevel Programming Problem, *Computers and Operations Research*, vol. 9, pp. 77-100
11. T. Başar and G.J. Olsder (1982) *Dynamic Noncooperative Game Theory*, Academic Press, London.
12. M. Beckmann, C.B. McGuire and C.B. Winsten (1956) *Studies in Economics of Transportation*, Yale University Press
13. O. Ben-Ayed, C.E. Blair (1990) Computational Difficulties of Bilevel Linear Programming, *Operations Research*, vol. 38, pp. 556-560
14. O. Ben-Ayed, D. E. Boyce and C.E. Blair (1988) A General Bilevel Linear Programming Formulation of the Network Design Problem, *Transportation Research*, vol. 22B, pp. 311-318

15. O. Ben-Ayed, C.E. Blair, D.E. Boyce and L.J. LeBlanc (1992) Construction of a Real-World Bilevel Linear Programming Model of the Highway Network Design Problem, In :: [5], pp. 219-254
16. J. Berechman (1984) Highway-Capacity Utilization and Investment in Transportation Corridors, *Environment and Planning*, vol. 16A, pp. 1475-1488
17. M. Bierlaire and P.L. Toint (1995) MEUSE: An Origin-Destination Matrix Estimator that Exploits Structure, *Transportation Research*, vol.29B, pp. 47-60
18. D.E. Boyce (1984) Urban Transportation Network-Equilibrium and Design Models: Recent Achievements and Future Prospects, *Environment and Planning*, vol. 16A, pp. 1445-1474
19. G.E. Cantarella and A. Sforza (1987) Methods for Equilibrium Network Traffic Signal Setting, In :: [75], pp. 69-89
20. E. Cascetta and S. Nguyen (1988) A Unified Framework for Estimating or Updating Origin-Destination Matrices from Traffic Counts, *Transportation Research*, vol. 22B, pp. 437-455
21. Y. Chen and M. Florian (1991) The Nonlinear Bilevel Programming Problem - A General Formulation and Optimality Conditions, Centre de recherche sur les transports, Publication No.794, Université de Montréal, C.P.6128, Montréal, Québec, H3C3J7, Canada
22. S. Dafermos (1980) Traffic Equilibrium and Variational Inequalitie, *Transportation Science*, vol. 14, pp. 42-54
23. S. Dafermos (1982) The General Multimodal Network Equilibrium Problem with Elastic Demand, *Networks*, vol. 12, pp. 57-72
24. S. Dafermos and F.T. Sparrow (1969) The Traffic Assignment Problem for a General Network, *Journal of Research of the National Bureau of Standards*, vol. 73B, pp. 91-118
25. C. Daganzo (1982) Unconstrained Extremal Formulations of Some Transportation Equilibrium Problems, *Transportation Science*, vol 16, pp. 332-361
26. O. Damberg and A. Migdalas (1995) Efficient Solution of Traffic Assignment Problems with a Distributed Simplicial Decomposition Algorithm, (*To appear*)
27. G.A. Davis(1994) Exact Local Solution of the Continuous Network Design Problem via Stochastic User Equilibrium Assignment, *Transportation Research*, vol. 28B, pp. 61-75
28. R.S. Dembo and U. Tulowitzki (1988) Computing Equilibria on Large Multicommodity Networks: An Application of Truncated Quadratic Programming Algorithms, *Networks*, vol. 18, pp. 273-284
29. A.H. deSilva and G.P. McCormick (1992) Implicit Defined Optimization Problems, In :: [5], pp. 107-124
30. A.V. Fiacco (1976) Sensitivity Analysis for Nonlinear Programming Using Penalty Methods, *Mathematical Programming*, vol.10, pp. 287-311
31. A.V. Fiacco (1983) *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, Academic Press, New York
32. C.S. Fisk (1984) Game Theory and Transportation Systems Modelling, *Transportation Research*, vol. 18B, pp. 301-313
33. C.S. Fisk (1984) Optimal Signal Controls on Congested Networks, *Ninth International Symposium on Transportation and Traffic Theory*, VNU Science Press, pp. 197-216
34. C.S. Fisk (1986) A Conceptual Framework for Optimal Transportatiion Systems Planning with Integrated Supply and Demand Models, *Transportation Science*, vol. 20, pp. 37-47
35. C.S. Fisk (1988) On Combining Maximum Entropy Trip Matrix Estimation with User Assignment, *Transportation Research*, vol. 22B, pp. 69-73
36. C.S. Fisk (1989) Trip Matrix Estimation from Link Traffic Counts: The Congested Network Case, *Transportation Research*, vol. 23B, pp. 331-336
37. C.S. Fisk and D. Boyce (1983) A Note on Trip Matrix Estimation from Link Traffic Count Data, *Transportation Research*, vol. 17B, pp. 245-250
38. C.S. Fisk and S. Nguyen (1982) Solution Algorithms for Network Equilibrium Models with Assymetric User Costs, *Transportation Science*, vol. 16, pp. 361-381
39. M. Florian and Y. Chen (1991) A Bilevel Programming Approach to Estimating O-D Matrix by Traffic Counts, Centre de recherche sur les transports, Publication No.750, Université de Montréal, C.P.6128, Montréal, Québec, H3C3J7, Canada

40. M. Florian and Y. Chen (1992) A Successive Linear Approximation Method for the Bilevel O-D Matrix Adjustment Problem, Centre de recherche sur les transports, Publication No.807, Université de Montréal, C.P.6128, Montréal, Québec, H3C3J7, Canada
41. M. Florian and Y. Chen (1993) A Coordinate Descent Method for the Bilevel O-D Matrix Adjustment Problem, Centre de recherche sur les transports, Publication No.807, Université de Montréal, C.P.6128, Montréal, Québec, H3C3J7, Canada. (Paper presented at the IFORS Conference in Lisbon, Portugal, July 1993)
42. T. Friesz (1985) Transportation Network Equilibrium, Design and Aggregation: Key Development and Research Opportunities, *Transportation Research*, vol. 19A, pp. 413-427
43. T.L. Friesz and P.T. Harker (1985) Properties of the Iterative Optimization-Equilibrium Algorithm, *Civil Engineering System*, vol. 2, pp. 142-154
44. T.L. Friesz, R.L. Tobin and T. Miller (1988) Algorithms for Spatially Competitive Network Facility-Location, *Environment and Planning*, vol. 15B, pp. 191-203
45. T.L. Friesz, R.L. Tobin, H. Cho and N.J. Mehta (1990) Sensitivity Analysis Based Heuristic Algorithms for the Mathematics Programs with Variational Inequality Constraints, *Mathematical Programming*, vol. 48, pp. 265-284
46. T.L. Friesz, H.-J. Cho, N.J. Mehta, R.L. Tobin and G. Anandalingam (1992) A Simulated Annealing Approach to the Network Design Problem with Variational Inequality Constraints, *Transportation Science*, vol. 26, pp. 18-26
47. N.H. Gartner, S.B. Gershwin, J.D.C. Little and P. Ross (1980), Pilot Study of Computer-Based Urban Traffic Management, *Transportation Research*, vol. 14B, pp. 203-217
48. P. Hansen, B. Jaumard and G. Savard (1992) New Branch-and-Bound Rules for Linear Bilevel Programming, *SIAM Journal on Scientific and Statistical Computing*, vol. 13, pp. 1194-1217
49. P.T. Harker and T.L. Friesz (1984) Bounding the Solution of the Continuous Equilibrium Network Design Problem, In :: Ninth International Symposium on Transportation and Traffic Theory, VNU Science Press, pp. 233-252
50. A. Haurie and P. Marcotte (1986) A Game Theoretic Approach to Network Equilibrium, *Mathematical Programming Study*, vol. 26, pp. 252-255
51. D.W. Hearn, S. Lawphongpanich and J.A. Ventura (1985) Finiteness in Restricted Simplicial Decomposition, *Operations Research Letters*, vol. 4, pp.125-130
52. G. Improta (1987) Mathematical Programming Methods for Urban Network Control, In :: [75], pp. 35-68
53. R. Jeroslow (1985) The Polynomial Hierarchy and a Simple Model for Competitive Analysis, *Mathematical Programming*, vol. 32, pp. 146-164
54. W. Ködel (1969) *Graphentheoretische Methoden und ihre Anwendungen*, Springer-Verlag, Berlin, pp. 56-59
55. C.D. Kolstad (1985) A Review of the Literature on Bilevel Mathematical Programming. Los Alamos National Laboratory, Report LA-10284-MS, Los Alamos, NM.
56. C.D. Kolstad and L.S. Lasdon (1990) Derivative Evaluation and Computational Experience with Large Bilevel Mathematical Programming, *Journal of Optimization and Applications*, vol. 65, pp. 485-499
57. T. Larsson and M. Patriksson (1992) Simplicial Decomposition with Disaggregated Representation for the Traffic Assignment Problem, *Transportation Science*, vol. 26, pp. 4-17
58. T. Larsson, A. Migdalas and M. Patriksson (1993) The Application of Partial Linearization Algorithm to the Traffic Assignment Problem, *Optimization*, vol. 28, pp. 47-61
59. S. Lawphongpanich and D.W. Hearn (1984) Simplicial Decomposition of the Asymmetric Traffic Assignment Problem, *Transportation Research*, vol. 18B, pp. 123-133
60. L.J. LeBlanc (1975) An Algorithm for the Discrete Network Design Problem, *Transportation Science*, vol. 9, pp. 183-199
61. L.J. LeBlanc and M. Abdulaal (1984) A Comparison of User-Optimum Versus System-Optimum Traffic Assignment in Transportation Network Design, *Transportation Research*, vol. 18B, pp. 115-121
62. L.J. LeBlanc, R.V. Helgason and D.E. Boyce (1985) Improved Efficiency of the Frank-Wolfe Algorithm for Convex Network Programs, *Transportation Science*, vol. 19, pp. 445-462

63. L.J. LeBlanc and D.E. Boyce (1986) A Bilevel Programming Algorithm for Exact Solution of the Network Design Problem with User Optimal Flows, *Transportation Research*, vol. 20B, pp. 259-265
64. P. Marcotte (1983) Network Optimization with Continuous Control Parameters, *Transportation Science*, vol. 17, pp.181-197
65. P. Marcotte (1986) Network Design Problem with Congestion Effects: A Case of Bilevel Programming, *Mathematical Programming*, vol. 34, pp. 142-162
66. P. Marcotte (1988) A Note on a Bilevel Programming Algorithm by LeBlanc and Boyce, *Transportation Research*, vol. 22B, pp. 233-237
67. P. Marcotte and G. Marquis (1992) Efficient Implementation of Heuristics for the Continuous Network Design Problem, In :: [5], pp. 163-176
68. A. Migdalas (1994) A Regularization of the Frank-Wolfe Method and Unification of Certain Nonlinear Programming Methods, *Mathematical Programming*, vol. 56, 331-345
69. A. Migdalas (1995) When is Stackelberg Equilibrium Pareto Optimum?, In :: P. Pardalos, et al (ed.s) *Advances in Multicriteria Analysis*, Kluwer Academic
70. A. Migdalas and H. Tuy (1995) On the Bilevel Min Norm Problem, (*To appear*)
71. M. Minoux (1986) *Mathematical Programming - Theory and Algorithms*, Translated from the Frence by S. Vajda, John Wiley & Sons, Chichester
72. S. Nguyen (1977) Estimation of an O-D Matrix from Network Data - A Network Equilibrium Approach, Centre de recherche sur les transports, Publication No.60, Université de Montréal, C.P.6128, Montréal, Québec, H3C3J7, Canada
73. S. Nguyen (1983) Inferring Origin-Destination Demands from Network Data, Associaziione Italiana di Recerca Operativa, Atti delle Giomate di Lavoro 1983, Napoli - Castel dell' Oro, 26-28 Settembre 1983,(*see also [74]*)
74. S. Nguyen (1984) Estimating Origin-Destination Matrices from Observed Flows, In *Transportation Planning Models*, M. Florian (Ed.), pp. 363-380
75. A.R. Odoni, L. Bianco and G. Szegö (1987) (Ed.s) *Flow Control of Congested Networks*, NATO ASI Series, Series F: Computer and Systems Sciences, vol. 38, Springer-Verlag, Berlin.
76. M. Patriksson (1994) *The Traffic Assignment Problem - Models and Methods*, VSP, Utrecht, The Netherlands
77. H. Poorzahedy and M.A. Turnquist (1982) Approximate Algorithms for the Discrete Network Design, *Transportation Research*, vol. 16B, pp. 45-55
78. M. Simaan and J.B. Cruz, Jr (1973) On the Stackelberg Strategy in Nonzero-Sum Games, *Journal of Optimization Theory and Applications*, vol 11, pp. 533-555
79. Y. Sheffi (1985) Urban Transportation Networks, Prentice Hall, Englewood Cliffs, NJ
80. M.J. Smith (1979) Existence, Uniqueness and Stability of Traffic Equilibria, *Transportation Research*, vol. 1B, pp. 295-304
81. H. Spiess (1990) A Gradient Approach for the O-D Matrix Adjustment Problem, Centre de recherche sur les transports, Publication No.693, Université de Montréal, C.P.6128, Montréal, Québec, H3C3J7, Canada
82. H. von Stackelberg (1952) *The Theory of the Market Economy*, Oxford University Press.
83. P.A. Steenbrink (1974) Transport Network Optimization in the Dutch Integral Transportation Study, *Transportation Research*, vol. 8, pp. 11-27
84. A. Steenbrink (1974) *Optimization of Transport Networks*, John Wiley & Sons, London
85. S. Suh and T.J. Kim (1992) Solving Nonlinear Programming Models of the Equilibrium Network Design Problem : A Comparative Review, In :: [5], pp. 203-218
86. C. Suwansirikul and T.L. Friesz (1987) Equilibrium Decomposed Optimization : A Heuristic for the Continuous Equilibrium Network Design Problem, *Transportation Science*, vol. 21, pp. 254-263
87. H. Tan, S.B. Gerashwin and M. Athans (1979) Hybrid Optimization in Urban Traffic Networks, Report DOT-TSC-RSP-79-7, MIT, Cambridge, Mass.
88. A.N. Tikhonov and V.Y. Arsenin (1977) *Solution of Ill-Posed Problems*, Translated from the Russian by F. John, John Wiley & Sons, New York.
89. R.L. Tobin (1986) Sensitivity Analysis for Variational Inequalities, *Journal of Optimization Theory and Applications*, vol. 48, pp. 191-204

90. R.L. Tobin and T.L. Friesz (1988) Sensitivity Analysis for Equilibrium Network Flows, *Transportation Science*, vol. 22, pp. 242-250
91. H. Tuy, A. Migdalas and P. Värbrand (1993) A Global Optimization Approach for the Linear Two-level Program, *Journal of Global Optimization*, vol. 3, pp. 1-23
92. H. Tuy, A. Migdalas and P. Värbrand (1994) A Quasiconcave Minimization Method for Solving Linear Two-Level Programs, *Journal of Global Optimization*, vol. 4, pp. 243-263
93. G.Ünlu (1987) A Linear Bilevel Programming Algorithm Based on Bicriteria Programming, *Computers and Operations Research*, vol. 14, pp. 173-179
94. J.H. Van Zuylen and L.G. Willumsen (1980) The Most Likely Trip Matrix Estimation from Traffic Counts, *Transportation Research*, vol. 14B, pp. 281-293
95. L.N. Vicente and P.H. Calamai (1994) Bilevel and Multilevel Programming: A Bibliographic Review, *Journal of Global Optimization*, vol. 5, pp. 291-306
96. L. Vicente, G. Savard and J. Júdice (1994) Descent Approaches for Quadratic Bilevel Programming, *Journal of Optimization Theory and Applications*, vol. 81, pp. 379-399
97. J.G. Wardrop (1952) Some Theoretical Aspects of Road Traffic Research, *Proceedings of the Institute of Civil Engineering*, Part II, pp. 325-378
98. U.-P. Wen and S.-T. Hsu (1989) A Note on a Linear Bilevel Programming Algorithm Based on Bicreteria Programming, *Computers and Operations Research*, vol. 16, pp. 79-83
99. H. Yang and S. Yagar (1994) Traffic Assignment and Traffic Control in General Freeway-Arterial Corridor Systems, *Transportation Research*, vol. 28B, pp. 463-486
100. H. Yang and S. Yagar (1995) Traffic Assignment and Signal Control in Saturated Road Networks, *Transportation Research*, vol. 29A, pp. 125-139
101. H. Yang, T. Sasaki, Y. Iida and Y. Asakura (1992) Estimation of Origin-Destination Matrices from Link Traffic Counts on Congested Networks, *Transportation Research*, vol. 26B, pp. 417-434
102. H. Yang, Y.Iida and T. Sasaki (1984) The Equilibrium-Based Origin-Destination Matrix Estimation Problem, *Transportation Research*, vol. 28B, pp. 23-33
103. H. Yang, S. Yagar, Y. Iida and Y. Asakura (1994) An Algorithm for the Inflow Control Problem on Urban Freeway Networks with User-Optimal Flows, *Transportation Research*, vol. 28B, pp. 123-139
104. S. Zukhovitshki, R. Polyak and M. Primak (1971) Concave n-person Games (Numerical Methods), *Ekonom. i Mat. Metody*, vol. 7, pp. 888-900