Les Cahiers du GERAD ISSN: 0711–2440

A Bilevel Programming Approach to Optimal Price Setting

Patrice Marcotte
Gilles Savard

G–2000–66

December 2000

A Bilevel Programming Approach to Optimal Price Setting

Patrice Marcotte
DIRO and CRT
Université de Montréal
marcotte@iro.umontreal.ca

Gilles Savard
MAGI and GERAD
École Polytechnique de Montréal
gilles.savard@polymtl.ca

December, 2000

Les Cahiers du GERAD

G–2000–66

Abstract

In this paper, we survey applications and algorithms pertaining to an important class of price setting problems formulated in the framework of bilevel programming.

Keywords: Bilevel programming. Pricing. Revenue Management. Transportation.

Résumé

Plusieurs problèmes de tarification se formulent naturellement sous la forme d'un programme mathématique à deux niveaux. Ce rapport présente un survol de modèles de tarification formulés sur des réseaux ainsi que les algorithmes adaptés à leur résolution.

Introduction

Bilevel programming is the adequate framework for asymmetric games where one player (the *leader*) calls the shots first, taking into account the optimal reaction of the second player (the *follower*). If one denotes by x (respectively y) the decision vector of the leader (respectively the follower), a bilevel program can be expressed as

$$\begin{array}{ll} \min\limits_{x,y} & f(x,y) \\ \text{subject to} & (x,y) \in X \\ & y \in \mathcal{S}(x), \end{array}$$

where $\mathcal{S}(x)$ denotes the set of optimal solutions of a mathematical program parameterized in the leader's vector x, i.e.,

$$\begin{array}{rcll} \mathcal{S}(x) & = & \arg\min\limits_{y} & g(x,y) \\ & & \text{subject to} & (x,y) \in Y. \end{array}$$

The above formulation implicitly assumes that, if the lower level problem admits multiple solutions, ties are broken in favor of the leader. Alternative situations, where the follower reacts in an antagonistic fashion, have been analyzed by Loridan and Morgan [13].

Bilevel programs are closely related to mathematical programs with equilibrium constraints (MPECS), where the lower level corresponds to an equilibrium problem. Indeed, whenever the objective g of the lower level program is differentiable and convex in y and the set Y is convex, $y \in \mathcal{S}(x)$ if and only if $(x,y) \in Y$ and satisfies the variational inequality

$$\langle \nabla_y g(x,y), y - y' \rangle \leq 0$$

for all y' such that $(x,y') \in Y$. Letting $Y(x) = \{y : (x,y) \in Y\}$ this leads to the one-level optimization formulation

$$\begin{array}{lll} \min\limits_{x,y} & f(x,y) & \\ \text{subject to} & (x,y) \in X & \\ & y \in Y(x) & \\ & \langle \nabla_y g(x,y), y - y' \rangle \leq 0 & \forall y' \in Y(x) \end{array}$$

which subsumes the more general form of an MPEC:

$$\begin{array}{lll} \min\limits_{x,y} & f(x,y) & \\ \text{subject to} & (x,y) \in X & \\ & y \in Y(x) & \\ & \langle G(x,y), y - y' \rangle \leq 0 & \forall y' \in Y(x), \end{array}$$

where the vector function G need not be a gradient mapping with respect to the variable y. Conversely, an MPEC can be reformulated as a standard bilevel program by noting that a vector y is solution of the lower level variational inequality if and only if it globally minimizes, with respect to the argument y, the strongly convex function $\text{gap}(x, y)$ defined as (see Fukushima [5]):

$$\text{gap}(x, y) = \max_{y' \in Y(x)} \langle G(x, y), y - y' \rangle - \frac{1}{2}\|y - y'\|^2.$$

The reader interested in the theory and applications of bilevel programming and MPEC is referred to the recent books by Shimizu, Ishizuka and Bard [16] and by Luo, Pang and Ralph [14].

Being generically nonconvex and nonsmooth, bilevel programs are difficult optimization problems. Even in the simple situation where both objectives are affine and the constraint sets are polyhedral, determining whether a solution is locally optimal is strongly NP-hard. This explains why global optimization techniques such as implicit enumeration, cutting planes or metaheuristics have been proposed for its solution (see e.g. [6] and [7]). These are most successful when the set $\mathcal{S}(x)$ assumes a piecewise polyhedral structure, for instance when the lower level problem takes the form of a convex quadratic program. In the absence of such property, two main lines of attack have been pursued. The first, based on sensitivity analysis, adapts descent methods that are compatible with the optimality requirements of the follower, and relies on recent nonsmooth analysis results; the work of Kocv̌ara, Outrata and Zowe [15] is typical of this trend. A drawback of this approach is that, even under strong regularity assumption, it may fail to uncover even a local optimum for the bilevel program.

A second approach applies standard optimization techniques to a one-level reformulation, smooth or not, of the bilevel program. A recent member of this family have been proposed by Scholtes and Stohr [17].

The present paper focuses on a specific class of bilevel problems that arises naturally when tariffs, tolls or devious taxes are imposed on a set of commodities. Not only does this class encompass several important optimization problems encountered in the transportation, telecommunication and airline industries, but its structure makes it amenable to efficient solution techniques. In this paper we present several such models and briefly discuss, in the final section, algorithmic approaches, either exact or heuristic, that can be applied to large scale problems within this class. Throughout the paper, the words “price” and “tax” are synonymous.

The price setting problem

Let x and y be real vectors that specify the respective levels of taxed and untaxed activities (commodities or services), and T be a tax vector attached to the activity vector x. For a given vector T, in control of the leader, the follower strives to minimize its operating costs, while the leader seeks to maximize the revenues raised from taxation. If one denote by F and f the leader's and follower's respective objective functions, the leader maximizes his profit by solving the bilevel program

$$\begin{array}{ll} \max\limits_{T,x,y} & F(T,x,y) \\ \text{subject to} & (x,y) \in \arg \min\limits_{(x',y')\in\Pi} f(T,x',y') \end{array} \tag{1}$$

where Π represents the constraint set of the second level player. From now on, we will record programs of the form (1) in the vertical format:

$$\begin{array}{ll} \max\limits_{T} & F(T,x,y) \\ \min\limits_{x,y} & f(T,x,y) \\ \text{subject to} & (x,y) \in \Pi. \end{array} \tag{2}$$

This seemingly simplistic model can cover a wide variety of situations. For instance the vector T may embody subsidies as well as taxes, while the vectors x and y may represent consumption or production levels. Alternatively, the lower level can represent the group behavior of economic agents competing for scarce resources; if the equilibrium states of the system are the solutions of a variational inequality parameterized in the leader's decision variables, we obtain an MPEC.

Let us first consider a basic model where the leader's revenues are proportional to tax and consumption levels, and where all constraints are linear. The resulting bilevel program, where both objectives are bilinear, takes the form

$$\begin{array}{ll} \max\limits_{T} & Tx \\ \min\limits_{x,y} & (c_1+T)x + c_2 y \\ \text{subject to} & A_1x + A_2y = b \\ & x, y \geq 0. \end{array} \tag{3}$$

From the leader's perspective, the objective function Tx is discontinuous at the points T that induce a change of optimal basis in the follower's linear program.

Assuming that the polyhedron $\{(x,y) : A_1x + A_2y = b, x, y \geq 0\}$ is bounded and that the recourse polyhedron $\{y : A_2y = b, y \geq 0\}$ is nonempty, the lower level always admits an optimal solution for any value of the tax vector T. Therefore, one can replace the lower level problem by its primal-dual optimality conditions to obtain the mathematical program with linear and complementarity constraints

$$\begin{array}{ll} \max\limits_{T,x,y,\lambda} & Tx \\ & \\ \text{subject to} & A_1x + A_2y = b \\ & x, y \geq 0 \\ & \lambda A_1 \leq c_1 + T \\ & \lambda A_2 \leq c_2 \\ & (c_1 + T - \lambda A_1)x = 0 \\ & (c_2 - \lambda A_2)y = 0 \end{array}$$

or the equivalent program

$$\begin{array}{ll} \max\limits_{T,x,y,\lambda} & \lambda b - (c_1 x + c_2 y) \\ \\ \text{subject to} & A_1 x + A_2 y = b \\ & x, y \geq 0 \\ & \lambda A_1 \leq c_1 + T \\ & \lambda A_2 \leq c_2 \\ & (c_1 + T - \lambda A_1) x = 0 \\ & (c_2 - \lambda A_2) y = 0. \end{array} \tag{4}$$

It is not difficult to see that (4) admits an optimal solution of the form $T = \lambda A_1 - c_1$. Upon substitution, we obtain the simplified model

$$\begin{array}{llcl} \max\limits_{x,y,\lambda} & \lambda b & - & (c_1 x + c_2 y) \\ \text{subject to} & \lambda A_2 \leq c_2 & & A_1 x + A_2 y = b \\ & & & x, y \geq 0 \\ & & \boxed{(c_2 - \lambda A_2) y = 0} & \end{array} \tag{5}$$

If we relax its complementarity constraint, (5) decomposes into two linear programs involving respectively the dual vector λ and the primal vectors x and y. The linear program associated with the primal variables corresponds to the solution of the lower level problem of (3) with taxes set at zero. The dual of the linear program associated with the dual variables is again the lower level linear program of (3), where the choice of activities is restricted to untaxed ones or, equivalently, where taxes are set to arbitrary high values.

Returning to the single-level formulation (5), one can penalize the complementarity constraint into the objective function, yielding a bilinear problem separable in the dual vector λ on the one hand, and in the primal vectors x and y on the other hand:

$$\begin{array}{ll} \max\limits_{x,y,\lambda} & \lambda b - (c_1 x + c_2 y) - M(c_2 - \lambda A_2) y \\ \\ \text{subject to} & \lambda A_2 \leq c_2 \\ & A_1 x + A_2 y = b \\ & x, y \geq 0. \end{array} \tag{6}$$

Labbé, Marcotte and Savard [11] have established the existence of an exact penalty parameter M^* such that any optimal solution of (6) is also optimal for (5) whenever M exceeds M^*. Now, for fixed primal variables x and y, let us replace the objective function λb by its dual objective; this yields the linear bilevel program:

$$\begin{aligned} \max_{x,y} \quad & -c_1x - (M+1)c_2y + c_2y' \\ \text{subject to} \quad & A_1x + A_2y = b \\ & x, y \geq 0 \\ \\ \min_{y'} \quad & c_2y' \\ \text{subject to} \quad & A_2y' = b + MA_2y \\ & y' \geq 0. \end{aligned} \tag{7}$$

An intuitive economic interpretation of the linear bilevel program (7) in terms of a *second best* alternative for the follower has been discussed [11].

We close this section with computational complexity considerations. In the absence of constraints on the tax vector T, we could not prove that (3) is NP-hard, although we suspect it is. However we proved that a variant of (3) where taxes are bounded from below is strongly NP-hard. The proof relies on a reduction from the "Hamiltonian Path" problem in a directed graph to a price setting problem with lower bound constraints.

Toll setting

The problem of selecting optimal highway tolls clearly fits our price-setting framework. Let us consider a multicommodity network where each commodity $k \in \mathcal{K}$ is associated with an origin-destination pair $(o(k), d(k))$ of a transportation network G defined by its node set $\mathcal{N}$ and arc set $\mathcal{A}$. The set $\mathcal{A}$ is partitioned into the subset $\mathcal{A}_1$ of toll arcs and the subset $\mathcal{A}_2$ of toll-free arcs. With each arc a of $\mathcal{A}_1$ is associated a generalized travel cost composed of a fixed part c_a representing the minimal travel cost per unit and an additional toll T_a, converted into time units. Any toll-free arc a of $\mathcal{A}_2$ bears a fixed unit travel cost d_a and we assume that the toll T_a cannot exceed a prescribed upper bound $T_a^{\max}$, which could be infinite.

A travel demand vector $\{n_k\}_{k\in\mathcal{K}}$ induces the nodal demand vectors

$$b_i^k = \begin{cases} n^k & \text{if } i = o(k), \\ -n^k & \text{if } i = d(k), \\ 0 & \text{otherwise.} \end{cases}$$

The lower level variable x_a^k corresponds to the number of users of commodity k on arc $a \in \mathcal{A}_1$ and the variable y_a^k to the number of users of commodity k on arc $a \in \mathcal{A}_2$.

Neglecting congestion effects, assuming that demand is fixed and that users minimize their individual generalized travel costs, the toll setting problem can be formulated as a bilevel program with bilinear objectives and linear constraints:

$$\begin{aligned}
&\max_{T} \sum_{a\in\mathcal{A}_1} T_a \sum_{k\in\mathcal{K}} x_a^k && \\
\text{subject to}\quad & T_a \le T_a^{\max} && \forall a \in \mathcal{A}_1 \\
&\min_{x,y} \sum_{k\in\mathcal{K}} \Big(\sum_{a\in\mathcal{A}_1} (c_a + T_a) x_a^k + \sum_{a\in\mathcal{A}_2} d_a y_a^k \Big) && \\
\text{subject to} & \sum_{a\in i^+} (x_a^k + y_a^k) - \sum_{a\in i^-} (x_a^k + y_a^k) = b_i^k && \forall k\in\mathcal{K} \quad \forall i \in \mathcal{N} \\
& x_a^k \ge 0 && \forall k\in\mathcal{K} \quad \forall a \in \mathcal{A}_1 \\
& y_a^k \ge 0 && \forall k\in\mathcal{K} \quad \forall a \in \mathcal{A}_2,
\end{aligned}$$

where i^+ (respectively i^-) denotes the set of arcs having i as their tail node (respectively head node).

One can introduce more realism into the previous model by considering congestion effects and/or a nonuniform distribution of the "trade-off value of time" across the population. These features will be incorporated in the next applications.

Price setting of telecommunication networks

This application concerns the optimal pricing of links in a packet-switched telecommunication network. We assume that the users select a telecommunication provider according to two key criteria: cost and quality of service, the latter being in direct relationship with the capacities of the links. The problem of the leader company is to price the arcs of its subnetwork such as to maximize profit, while taking into account the user-optimized behavior of the customers. If the arrival rate of messages follows a Poisson process and is independent of the service time, and the length of a message is distributed according to an exponential random variable, then the average delay of a message through the network is given by the formula

$$T(f) = \frac{1}{\gamma} \sum_{a\in\mathcal{A}} \left(\frac{f_a}{C_a - f_a} + \mu f_a p_a \right), \tag{8}$$

where f_a denotes the total flow on arc $a \in \mathcal{A}$, γ the total demand on the network, C_a the capacity of the arc $a \in \mathcal{A}$, μ the average message length and p_a the propagation delay along arc $a \in \mathcal{A}$. The first term in the summation, the node delay, reflects congestion at nodes. Under normal conditions, the second term (the arc delay) is negligible and can safely be discarded.

Upon introduction of trade-off parameters α_a that translate quality of service in terms of cost units, and of average generalized costs d_a^k along the arcs of the competing firms, the price-setting problem takes the form

$$\begin{aligned}
&\max_{T} && \sum_{a\in\mathcal{A}_1}\sum_{k\in\mathcal{K}} T_a^k x_a^k && \\
&\text{subject to} && T_a \le T_a^{\max} && \forall a\in\mathcal{A}_1 \\
& && && \\
&\min_{f,x,y} && \sum_{a\in\mathcal{A}} \frac{\alpha_a f_a}{\gamma(C_a - f_a)} + \sum_{k\in\mathcal{K}}\sum_{a\in\mathcal{A}_1} T_a^k x_a^k + \sum_{k\in\mathcal{K}}\sum_{a\in\mathcal{A}_2} d_a^k y_a^k && \\
&\text{subject to} && \sum_{a\in i^+}(x_a^k + y_a^k) - \sum_{a\in i^-}(x_a^k + y_a^k) = b_i^k && \forall k\in\mathcal{K}\quad \forall i\in\mathcal{N} \\
& && f_a = \sum_{k\in\mathcal{K}} x_a^k && \forall a\in\mathcal{A}_1 \\
& && f_a = \sum_{k\in\mathcal{K}} y_a^k && \forall a\in\mathcal{A}_2 \\
& && x_a^k \ge 0 && \forall k\in\mathcal{K}\quad \forall a\in\mathcal{A}_1 \\
& && y_a^k \ge 0 && \forall k\in\mathcal{K}\quad \forall a\in\mathcal{A}_2.
\end{aligned}$$

In this formulation, arc prices are not required to be identical for all commodities. This opens room for discrimination of the users and would make the problem entirely separable by commodity, hence trivial to solve, were it not for the congestion effects.

Yield management in the airline industry

Yield management in the airline industry has been an area of active research for the past few years (see [18]). It addresses four issues that deeply impact the industry revenues: forecasting, overbooking, seat allocation and pricing. Ideally, these four components should be part of an integrated profit-maximizing model. However, the complexity and the size of such a model would prevent the computer resolution of problem instances of any realistic size. At the present time, the four issues are treated independently, even though the strong interaction between seat allocation and pricing is widely acknowledged. Our model, which addresses jointly the issues of seat allocation and pricing, is distinguished by three key features: its bilevel nature allows for an endogenous representation of the price-demand relationship, network interactions among the various carriers are explicitly considered, and the utility function of each user takes into account three criteria: fare, time and quality of service. Two parameters, α and β, translate time and quality of service into cost units; these parameters are distributed across the population according to a density $\phi^k(\alpha,\beta)$ which may vary by origin-destination pair k. This flexibility allows the model to assign a given user to a fare class according to his personal preference.

Let T denote the fare vector and x (respectively y) denote the passenger flow vector for the leader airline (respectively the competition), indexed by booking classes b and flights f connecting the origin-destination pair k. The revenue of the leader airline is

$$f(T,x,y) = \sum_k \sum_{f\in\mathcal{C}_1} \sum_{b\in\mathcal{B}(f)} T^k_{bf} \int_\alpha \int_\beta x^k_{bf}(\alpha,\beta)\, d\alpha\, d\beta,$$

where $\mathcal{C}_1$ is the set of flights of the leader company and $\mathcal{B}(f)$ is the set of booking classes available on flight f. A passenger whose utility parameters are α and β will select a flight that minimizes his generalized cost

$$\text{fare} + (\alpha \times \text{flight length}) + (\beta \times \text{flight quality}).$$

Hence, at the lower level, the passengers are assigned to the flights and booking classes that minimize the objective

$$\begin{aligned} g(x,y) = \quad & \sum_k \Big\{ \sum_{f\in\mathcal{C}_1} \sum_{b\in\mathcal{B}(f)} \int_\alpha \int_\beta (T^k_{bf} + \alpha l_f + \beta q_{bf}) x^k_{bf}(\alpha,\beta)\, d\alpha\, d\beta + \\ & \sum_{f\in\mathcal{C}_2} \sum_{b\in\mathcal{B}(f)} \int_\alpha \int_\beta (T^k_{bf} + \alpha l_f + \beta q_{bf}) y^k_{bf}(\alpha,\beta)\, d\alpha\, d\beta \Big\}, \end{aligned}$$

where $\mathcal{C}_2$ is the set of flights of the competing firms and, l_f and q_{bf} the respective length and quality index associated with flight f and booking class b. The functions $x^k_{bf}(\alpha,\beta)$ and $y^k_{bf}(\alpha,\beta)$ are the flow densities for booking class b of flight k associated with parameter couples (α,β). Their integrals represent the respective market shares of the leader airline and the competition.

Let $F_1(k)$ (respectively $F_2(k)$) be the set of flights of the leader (respectively the competition) available for origin-destination k and $\{d_k\}_{k\in\mathcal{K}}$ the demand vector. Let $A(f)$ denote the set of legs that make up flight f. After incorporating the flow conservation and capacity constraints, we formulate a bilevel model involving both finite-dimensional (the fares) and infinite-dimensional (the flow densities) decision variables:

$$\begin{aligned} \max_{T,x,y} \quad & \sum_k \sum_{f\in\mathcal{C}_1} \sum_{b\in\mathcal{B}(f)} T^k_{bf} \int_\alpha \int_\beta x^k_{bf}(\alpha,\beta)\, d\alpha\, d\beta \\ \text{subject to} \quad & T^{\min}_{bf} \le T_{bf} \le T^{\max}_{bf} \qquad \forall f \in \mathcal{C}_1 \forall b \in \mathcal{B}(f) \\ \\ \min_{x,y} \quad & \sum_k \Big\{ \sum_{f\in\mathcal{C}_1} \sum_{b\in\mathcal{B}(f)} \int_\alpha \int_\beta (T^k_{bf} + \alpha l_f + \beta q_{bf}) x^k_{bf}(\alpha,\beta)\, d\alpha\, d\beta \\ & \quad + \sum_{f\in\mathcal{C}_2} \sum_{b\in\mathcal{B}(f)} \int_\alpha \int_\beta (T^k_{bf} + \alpha l_f + \beta q_{bf}) y^k_{bf}(\alpha,\beta)\, d\alpha\, d\beta \Big\} \\ \text{subject to} \quad & x_a = \sum_k \sum_{f:a\in A(f)} \sum_{b\in\mathcal{B}(f)} \int_\alpha \int_\beta x^k_{bf}(\alpha,\beta)\, d\alpha\, d\beta \qquad \forall a \in \mathcal{A}_1 \end{aligned}$$

$$\begin{aligned}
&\phi^k(\alpha,\beta)d^k = \sum_{b\in\mathcal{B}(f)}\Big\{\sum_{f\in F(k)} x^k_{bf}(\alpha,\beta) + \sum_{f\in F'(k)} y^k_{bf}(\alpha,\beta)\Big\} && && \forall k \in K\\
&x_a \le x_a^{\max} && && \forall a\in\mathcal{A}_1\\
&x^k_{bf}(\alpha,\beta) \ge 0 && & \forall f\in\mathcal{C}_1 & \forall b\in\mathcal{B}(f)\\
&y^k_{bf}(\alpha,\beta) \ge 0 && \forall k\in\mathcal{K} & \forall f\in\mathcal{C}_2 & \forall b\in\mathcal{B}(f).
\end{aligned}$$

Note that the above model is pessimistic in the sense that the competition's capacity is assumed to be illimited; this assumption makes sense if the leader company's market share is small.

Traffic management through link tolls

On a transportation network, Wardrop's user principle states that, at equilibrium, flows should be assigned to shortest routes with respect to current travel delays. If the network is heavily congested, such an assignment may lead to an inefficient use of the network capacity. It is well known that marginal cost pricing induces the network users to behave in a fashion that is socially optimal. However, this scheme is not alone to possess that property, and one might prefer to implement a regulating policy that minimizes the total amount of tolls raised from the users (see e.g. Hearn and Ramana [8] and Larsson and Patriksson [12]). Denoting by x^k the flow vector associated with origin-destination k, x the vector of aggregate flows, x^* the vector of system-optimal flows, $C_a(x)$ the congestion delay along arc a and $\{X_k\}_{k\in\mathcal{K}}$ the sets of feasible flows per commodity, this is achieved by solving the MPEC

$$\begin{aligned}
\min_{T\ge 0}\quad & \sum_{k\in K}\sum_{a\in\mathcal{A}} T_a x_a^{k*}\\
\text{subject to}\quad & \sum_{a\in\mathcal{A}}\Big(C_a(\sum_{l\in K} x_a^{l*}) + T_a\Big)(x_a^{k*} - x_a^k) \le 0 \quad \forall x^k \in X_k \quad \forall k\in K,
\end{aligned}$$

where $X_k = \{Bx^k = b^k, x^k \ge 0\}$.

Algorithms

The price setting problem

The basic bilinear model can be solved as a linear bilevel program using the reformulation (7). However, to preserve the structure of the problem, it is better to work directly on the formulation (5), which can be solved by a Branch-and-Bound procedure where branching is performed with respect to the complementarity constraint, i.e., the alternative

$$(c_2 - \lambda A_2)_i = 0 \quad \text{or} \quad y_i = 0.$$

Whenever the index set P of positive variables y_j $(j \in P)$ in an optimal solution is specified at a given node of the enumeration tree, an optimal λ-vector can be recovered by solving the linear program:

$$\begin{array}{rl} \max\limits_{\lambda} & \lambda b \\ \text{subject to} & (\lambda A_2 - c_2)_j \leq 0 \quad \forall j \notin P \\ & (\lambda A_2 - c_2)_j = 0 \quad \forall j \in P \\ & \lambda A_1 \leq c_1, \end{array}$$

whose dual

$$\begin{array}{rl} \min\limits_{x,y} & c_1 x_1 + c_2 y \\ \text{subject to} & A_1 x + A_2 y = b \\ & x \geq 0 \\ & y_j \quad \text{unrestricted } \forall j \in P \\ & y_j \geq 0 \quad \forall j \notin P \end{array}$$

has a structure similar to that of the original lower level LP. If this “inverse optimization” procedure is carried out at each node of the implicit enumeration tree, then the algorithm may heuristically be halted before termination and yet produce a solution of high quality. Actually, the dual vector λ derived from an optimal solution of the lower level problem with $T = 0$ frequently yields a very good initial solution.

In practice, it is frequently the case that the size of the taxed flow vector x is much smaller than the size of y. If this is the case, the knowledge of the vector x (not only the index set of its positive components, but also their values) allows to derive the values of y by solving the LP:

$$\begin{array}{rl} \min\limits_{y} & c_2 y \\ \text{subject to} & A_2 y = b - A_1 x \\ & y \geq 0. \end{array}$$

As before, an optimal tax vector corresponding to the vector y can easily be recovered.

Another approach, which is better suited to the solution of large scale problems, consists in penalizing the complementarity constraint of (5) to obtain the bilinear program

$$\begin{array}{rl} \min\limits_{x,y,\lambda} & c_1 x + c_2 y - \lambda b + M(c_2 - \lambda A_2) y \\ \text{subject to} & A_1 x + A_2 y = b \\ & x, y \geq 0 \\ & \lambda A_2 \leq c_2 \end{array} \tag{9}$$

for some penalty parameter M. Variants of the Gauss-Seidel iterative procedure, coupled to a clever update of the parameter M, performed very well on single-commodity toll setting problems.

Toll setting

In contrast with the basic model, the multicommodity network model has three distinguished features: (i) its network structure, (ii) upper bounds on the tolls and (iii) tolls that apply to total flow, not individual commodity flows. The presence of a network structure allows for a mixed-integer formulation that involves a relatively small number of integer variables, i.e., one per toll arc and per commodity. We express the flow variables as proportions of the demand n_k associated with the origin-destination couple k and the nodal demands are set to $e_i^k = \text{sgn}(b_i^k)$. We then obtain the formulation

$$\begin{array}{lll}
\max\limits_{T,x,y,\lambda} & \sum\limits_{k\in\mathcal{K}} \sum\limits_{a\in\mathcal{A}_1} n^k T_a^k & \\
\text{subject to} & \sum\limits_{a\in i^+} (x_a^k + y_a^k) - \sum\limits_{a\in i^-} (x_a^k + y_a^k) = e_i^k & \forall i \in \mathcal{N} \quad \forall k \in \mathcal{K} \\
& \lambda_i^k - \lambda_j^k \leq c_a + T_a & \forall a = (i,j) \in \mathcal{A}_1 \quad \forall k \in \mathcal{K} \\
& \lambda_i^k - \lambda_j^k \leq d_a & \forall a \in \mathcal{A}_2 \quad \forall k \in \mathcal{K} \\
& \sum\limits_{a\in\mathcal{A}_1} (c_a\, x_a^k + T_a^k) + \sum\limits_{a\in\mathcal{A}_2} d_a\, y_a^k = \lambda_{o(k)}^k - \lambda_{d(k)}^k & \forall k \in \mathcal{K} \\
& -Mx_a^k \leq T_a^k \leq Mx_a^k & \forall k \in \mathcal{K} \quad \forall a \in \mathcal{A}_1 \\
& -M(1-x_a^k) \leq T_a^k - T_a \leq M(1-x_a^k) & \forall k \in \mathcal{K} \quad \forall a \in \mathcal{A}_1 \\
& x_a^k \in \{0,1\} & \forall k \in \mathcal{K} \quad \forall a \in \mathcal{A}_1 \\
& y_a^k \geq 0 & \forall k \in \mathcal{K} \quad \forall a \in \mathcal{A}_2 \\
& T_a \leq T_a^{\max} & \forall a \in \mathcal{A}_1.
\end{array}$$

The fourth constraint of the above program ensures that complementarity slackness of the lower level problem is satisfied, while the next two constraints ensure that the commodity tolls T_a^k are equal to the actual common toll T_a if the commodity flow x_a^k is positive. This formulation allowed us to solve to optimality small instances of the multicommodity toll problem (see e.g. Brotcorne [2] and Brotcorne et al [1]) .

While the treatment of upper bounds in the multicommodity model constitutes a mere technicality, the presence of multiple commodities makes the resolution of (9) by a Gauss-Seidel strategy significantly more complex. Specifically, the inverse optimization procedure described earlier must now rely on the solution of a multicommodity flow problem with upper bounds on arc flows. To avoid this difficulty, we propose simply to replace the global toll vector T by commodity tolls T^k that must satisfy the compatibility constraint

$$T_k = T_1 \qquad \forall k.$$

By penalizing this constraint into the objective, one obtains the bilinear-quadratic program

$$\begin{array}{lll}\max\limits_{\lambda,x,y} & \multicolumn{2}{l}{\displaystyle\sum_{k\in K}\Big[\lambda^k b^k-(c_1x^k+c_2y^k)-M_1\|(\lambda^k-\lambda^1)A_1\|^2} \\ & \multicolumn{2}{r}{-M_2(c_2-\lambda^kA_2)y^k\Big]} \\ \text{subject to} & A_1x^k+A_2y^k=b^k & \forall k\in K \\ & x^k,y^k\geq 0 & \forall k\in K \\ & (c_2-\lambda^kA_2)y^k=0 & \forall k\in K \\ & \lambda^kA_1-c_1\leq u & \forall k\in K. \end{array}$$

Once the quadratic penalty term is linearized (à la Frank-Wolfe), the Gauss-Seidel strategy may be applied to the resulting bilinear program. Provided that the penalty parameters M_1 and M_2 are calibrated in a suitable way, the coupling of the Gauss-Seidel and inverse optimization strategies could uncover solutions of large-scale problems whose objective values were typically within one percent of the optimal values, when these were known (see[1]).

Price setting of telecommunication networks

The algorithmic approaches presented for the multicommodity network model can be extended to the nonlinear telecommunication model. To see this, let us replace the lower level by its optimality conditions:

$$\begin{array}{lll}\max\limits_{T,x,y,\lambda} & \displaystyle\sum_{a\in\mathcal{A}_1}\sum_{k\in\mathcal{K}}T_a^kx_a^k & \\ \text{subject to} & T_a\leq T_a^{\max} & \forall a\in\mathcal{A}_1 \\ & T_a^k+\dfrac{\alpha_aC_a}{\gamma(C_a-f_a)^2}-\lambda_j^k+\lambda_i^k-\mu_a^k=0 & \forall a\in\mathcal{A}_1\quad\forall k\in\mathcal{K} \\ & d_a^k+\dfrac{\alpha_aC_a}{\gamma(C_a-f_a)^2}-\lambda_j^k+\lambda_i^k-\mu_a^k=0 & \forall a\in\mathcal{A}_2\quad\forall k\in\mathcal{K} \\ & \displaystyle\sum_{a\in i^+}(x_a^k+y_a^k)-\sum_{a\in i^-}(x_a^k+y_a^k)=b_i^k & \forall i\in\mathcal{N}\quad\forall k\in\mathcal{K} \\ & f_a=\displaystyle\sum_{k\in\mathcal{K}}x_a^k & \forall a\in\mathcal{A}_1 \\ & f_a=\displaystyle\sum_{k\in\mathcal{K}}y_a^k & \forall a\in\mathcal{A}_2 \\ & \mu_a^kx_a^k=0,\quad \mu_a^k\geq 0,\quad x_a^k\geq 0 & \forall k\in\mathcal{K}\quad\forall a\in\mathcal{A}_1 \\ & \mu_a^ky_a^k=0,\quad \mu_a^k\geq 0,\quad y_a^k\geq 0 & \forall k\in\mathcal{K}\quad\forall a\in\mathcal{A}_2. \end{array}$$

Using the same argument as the one developed for the general model, we obtain, after straightforward albeit tedious calculations:

$$\begin{aligned}\sum_{a\in\mathcal{A}_1}\sum_{k\in\mathcal{K}} T_a^k x_a^k &= \sum_{i\in N}\sum_{k\in\mathcal{K}}\lambda_i^k b_i^k - \sum_{a\in\mathcal{A}_1}\frac{\alpha_a C_a f_a}{\gamma(C_a-f_a)^2}\\ &\quad -\sum_{a\in\mathcal{A}_2}\frac{\alpha_a C_a f_a}{\gamma(C_a-f_a)^2} - \sum_{a\in\mathcal{A}_2}\sum_{k\in\mathcal{K}} d_a^k f_a^k.\end{aligned}$$

We linearize both the congestion function (by a piecewise linear curve) and the complementarity constraints (as previously and with the help of the binary variables z_a^k). If m_{ap} corresponds to a breakpoint, w_{ap} to the binary variable associated with the pth segment and, t_{ap} the variable associated with the convex combination of the p^{th} segment, one obtains the mixed integer linear program

$$\begin{aligned}\max_{T,x,y,w,z}\quad & \sum_{i\in N}\sum_{k\in\mathcal{K}}\lambda_i^k b_i^k - \sum_{a\in\mathcal{A}_1}\sum_{p=1}^{n+1}\frac{\alpha_a C_a}{\gamma} t_{ap} h(m_{ap})\\ & -\sum_{a\in\mathcal{A}_2}\sum_{p=1}^{n+1}\frac{\alpha_a C_a}{\gamma} t_{ap} h(m_{ap})\\ & -\sum_{a\in\mathcal{A}_2}\sum_{k\in\mathcal{K}} d_a^k f_a^k\end{aligned}$$

subject to

$$\begin{aligned}
& T_a \le T_a^{\max} && \forall a\in\mathcal{A}_1\\
& \lambda_j^k - \lambda_i^k - \sum_{p=1}^{n+1}\frac{\alpha_a C_a}{\gamma} t_{ap} g(m_{ap}) \le d_a^k && \forall k\in\mathcal{K}\quad \forall a\in\mathcal{A}_2\\
& f_a = \sum_{p=1}^{n+1} t_{ap} m_{ap} && \forall a\in\mathcal{A}\\
& f_a = \sum_{k\in\mathcal{K}} x_a^k && \forall a\in\mathcal{A}_1\\
& f_a = \sum_{k\in\mathcal{K}} y_a^k && \forall a\in\mathcal{A}_2\\
& \sum_{p=1}^{n+1} t_{ap} = 1 && \forall a\in\mathcal{A}\\
& t_{a1}\le w_{a1} \qquad t_{a(n+1)} \le w_{an} && \forall a\in\mathcal{A}\\
& t_{ap} \le w_{a(p-1)} + w_{ap} && p=2,...,n\quad \forall a\in\mathcal{A}
\end{aligned}$$

$$
\begin{aligned}
&\sum_{p=1}^{n} w_{ap} = 1 && \forall a \in \mathcal{A}\\
&\sum_{a\in i^+}(x_a^k + y_a^k) - \sum_{a\in i^-}(x_a^k + y_a^k) = e_i^k && \forall i \in \mathcal{N} \forall k \in \mathcal{K}\\
&d_a^k + \sum_{p=1}^{n+1} \frac{\alpha_a C_a}{\gamma} t_{ap} g(m_{ap}) - \lambda_j^k + \lambda_i^k - M z_a^k \leq 0 && \forall a \in \mathcal{A}_2 \quad \forall k \in \mathcal{K}\\
&f_a^k \leq M(1 - z_a^k) && \forall a \in \mathcal{A}_2 \quad \forall k \in \mathcal{K}\\
&f_a^k \geq 0 && \forall a \in \mathcal{A} \quad \forall k \in \mathcal{K}\\
&w_{ap} \in \{0,1\} && p \in \{1,2,...,n\} \quad \forall a \in \mathcal{A}\\
&t_{ap} \geq 0 && p \in \{1,2,\ldots,n+1\} \quad \forall a \in \mathcal{A}\\
&z_a^k \in \{0,1\} && \forall a \in \mathcal{A}_2 \quad \forall k \in \mathcal{K}.
\end{aligned}
$$

The above model can only be solved to optimality for small instances. For larger instances, Julsain [10] implemented the following procedure: for fixed price levels, an equilibrium flow solution of the lower level problem traffic assignment problem is obtained; next, the optimal price schedule corresponding to this flow is determined by solving a linear inverse optimization problem. The process is iterated until no change is observed, at which point the best solution obtained by the inverse optimization procedure is retained.

Yield management in the airline industry

The yield management problem is by far the most ambitious model presented in this survey, involving the solution of an infinite-dimensional lower level problem. Working in infinite dimension could, surpisingly, prove an asset from the computational point of view, since the continuous distribution of the trade-off values α and β across the population smoothes out the lower level reaction to the upper level fares and makes the lower level solution unique. It is then conceivable to address the bilevel problem as a single-level differentiable optimization problem, whose objective can be computed by tedious but rather straightforward implicit derivation rules. Alternatively, one could discretize the distribution functions and apply the algorithmic ideas presented for the other applications. A mixed continuous-discrete approach which mixes both ideas is currently investigated.

Traffic management through link tolls

In contrast with the previous applications, the traffic management problem is not a *bona fide* bilevel program but actually an inverse optimization problem that can be solved easily. Let us first analyze the case of a single origin (single commodity), which can be written, using obvious vector and matrix notation, as

$$\begin{array}{ll} \max\limits_{T\geq 0} & Tx^* \\ \text{subject to} & \langle C(x^*)+T, x^*-x\rangle \leq 0 \qquad \forall x \in X = \{Bx=b, x\geq 0\}. \end{array}$$

Since the solution of the lower level variational inequality is known to be x^*, this vector is a solution of the primal-dual system

$$\begin{array}{rcl} Bx^* &=& b \\ x^* &\geq& 0 \\ \lambda B &\leq& C(x^*)+T \\ \langle C(x^*)+T-\lambda B, x^*\rangle &=& 0. \end{array} \tag{10}$$

If x_a^* is positive, the corresponding toll is set to $T_a = (\lambda B - C(x^*))_a$; otherwise, the toll T_a can be set to any value that exceeds the maximum equilibrium delay, which is known. To simplify the presentation, we assume that all components of the vector x^* are positive. The traffic management problem can then be reduced to the linear program

$$\begin{array}{ll} \min\limits_{\lambda} & \langle \lambda B - C(x^*), x^*\rangle \\ \text{subject to} & \lambda B - C(x^*) \geq 0, \end{array}$$

whose dual

$$\max_{x\in X} C(x^*)x$$

consists in finding a longest path tree in an acyclic network, and is solvable by a greedy algorithm in linear time (Dial[3]).

In the multicommodity case, let $\overline{x}_a$ denote the total flow on arc a. The equivalent of (10) is

$$\begin{array}{ll} Bx^{k*} = b^k & \forall k \in \mathcal{K} \\ x^{k*} \geq 0 & \forall k \in \mathcal{K} \\ \lambda^k B \leq C(\overline{x}^*) + T & \forall k \in \mathcal{K} \\ \sum\limits_{k\in K} \langle C(\overline{x}^*) + T - \lambda^k B, x^{k*}\rangle = 0. & \end{array}$$

The dual of the problem of maximizing $T\overline{x}^*$ subject to the last two constraints is

$$\begin{array}{ll} \max\limits_{x,\alpha} & \langle C(\overline{x}^*), \alpha\overline{x}^* - \overline{z}\rangle \\ \text{subject to} & \overline{z} = \sum\limits_{k\in\mathcal{K}} z^k \\ & \overline{z} \leq (1+\alpha)\overline{x}^* \\ & Bz^k = \alpha b^k,\ z^k \geq 0 \qquad \forall k \in \mathcal{K} \end{array}$$

which, after performing the scaling $z = \alpha x'$, becomes

$$\begin{aligned} \max_{\alpha} \quad \phi(\alpha) \quad := \quad \alpha \max_{x' \in \prod_k X_k} \quad & \langle C(\overline{x}^*), \overline{x}^* - \overline{x}' \rangle \\ \text{subject to} \quad & \overline{x}' \leq (1 + 1/\alpha)\, \overline{x}^*. \end{aligned} \tag{11}$$

The function ϕ is concave, increasing, piecewise linear, and its maximum is achieved for any sufficiently large value of the variable α. For such values of α, an optimal tax vector will be recovered from the dual vector λ associated with the capacity constraint (11) of a linear multicommodity flow problem. A slightly different multicommodity formulation involving a nonnegativity constraint on the sum of commodity flows was derived by Dial [4].

While the multicommodity problem is considerably more difficult to solve than finding a longest path tree in an acyclic network, it is nonetheless manageable, even for large scale networks. Actually, by penalizing the upper bound constraints, an approximate solution can be computed by solving a convex flow problem which is at least as easy to solve as the traffic assignment problem which had to be solved in order to determine the system-optimal flow vector x^* in the first place.

If only a subset of arcs is subject to tolls, then it might be impossible to induce a system-optimal flow pattern. If one adopts as maximization criterion the social surplus, then the problem becomes an authentic bilevel program to which the techniques presented earlier can be applied.

Conclusion

In this paper, we have presented a short survey of pricing situations modeled as bilevel programs, as well as several avenues for their numerical resolution. We firmly believe that this approach will gain in popularity both in the economic and mathematical programming communities, and that the day is not far where these models will be routinely solved for near-optimal solutions.

References

[1] Brotcorne, L., Labbé, M., Marcotte, P. and Savard, G., “ A Bilevel Model for Toll Optimization on a Multicommodity Transportation Network”, Rapport technique CRT00, 1999.

[2] Brotcorne, L., “Approches opérationnelles et stratégiques des problèmes de trafic routier”, Ph.D. thesis, Université Libre de Bruxelles, February 1998.

[3] Dial, R. B., “Minimal-revenue congestion pricing Part II: An efficient algorithm for the general case”, *Transportation Research B* **34** (2000) 645–665.

[4] Dial, R. B., “Minimal-revenue congestion pricing part I: A fast algorithm for the single-origin casse case”, *Transportation Research B* **33** (1999) 189–202.

[5] Fukushima, M., “Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems”, *Mathematical Programming* **53** (1992) 99–110.

[6] Gendreau, M., Marcotte, P. and Savard, G., "A Hybrid tabu-ascent algorithm for the linear bilevel programming problem", *Journal of Global Optimization* **8** (1996) 217–232.

[7] Hansen, P., Jaumard, B. and Savard, G., "New branch-and-bound rules for linear bilevel programming", *JOTA* **22** (1992) 1194–1217.

[8] Hearn, D.W. and Ramana, M.V., "Solving congestion toll pricing models", in: *Equilibrium and Advanced Transportation Modelling*, Patrice Marcotte and Sang Nguyen (eds.), Kluwer, Dordrecht, pp. 109-123, 1998.

[9] Kočvara, M. and Outrata, J.V., "A nonsmooth approach to optimization problems with equilibrium constraints", in *Complementarity and variational problems. State of the art*, M.C. Ferris and J.S. Pang (eds.), SIAM, Philadelphia (1997).

[10] Julsain, H., "Tarification dans les réseaux de télécommunications: une approche par programmation mathématique à deux niveaux", Mémoire de maîtrise, École Polytechnique de Montréal, 1999.

[11] Labbé, M., Marcotte, P. and Savard, G., "A bilevel model of taxation and its application to optimal highway pricing", *Management Science* **44** (1998) 1595–1607 .

[12] Larsson, T. and Patriksson, M., "Traffic management through link tolls - an approach utilizing side constrained traffic equilibrium models", in: *Equilibrium and Advanced Transportation Modelling*, Patrice Marcotte and Sang Nguyen (eds.), Kluwer, Dordrecht, pp. 125-151, 1998.

[13] Loridan, P. and Morgan, J., "ϵ-regularized two-level optimization problems: approximation and existence results", *Fifth French-German Conference on Optimization*, Lecture Notes in Mathematics 1405, Springer-Verlag (1989) 99–113.

[14] Luo, Z.Q., Pang, J.S. and Ralph, D., *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge (1996).

[15] Outrata, J.V., Kocvara, M. and Zowe, J., *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints.* Kluwer Academic Publishers, Dordrecht, (1998).

[16] Shimizu, K., Ishizuka, Y, and Bard, J.F., *Nondifferentiable and Two-Level Mathematical Programming*, Kluwer, Boston (1997).

[17] Scholtes, S. and Stohr, M., "Exact penalization of mathematical programs with equilibrium constraints" *SIAM J. Control and Optimization* **37**, pp.617-652, (1999).

[18] Special issue on Yield Management, *Transportation Science* **33** (1999).