Chapter 7
Applications to Other Energy Systems

7.1 Consistent Conjectural Variations Equilibrium in a Mixed Oligopoly in Electricity Markets

Results described in this section are based mainly upon paper of Kalashnikov et al. [165], which also included applications to an oligopolistic market of electricity. Even if the main models and tools developed in the paper are not directly related to Bilevel Programming, they can be used to construct more complicated schemes involving the Stackelberg equilibrium and other bilevel-type concepts.

In more detail, this section deals with a model of mixed oligopoly with *conjectured variations equilibrium* (CVE). The agents' conjectures concern the price variations depending upon their production output's increase or decrease. We establish existence and uniqueness results for the conjectured variations equilibrium (called an *exterior equilibrium*) for any set of feasible conjectures. In order to introduce the notion of an *interior equilibrium*, we develop a *consistency criterion* for the conjectures (referred to as *influence coefficients*) and prove the existence theorem for the interior equilibrium (understood as a CVE with consistent conjectures). To prepare the base for the extension of our results to the case of non-differentiable demand functions, we also investigate the behavior of the consistent conjectures depending upon a parameter that represents the demand function's derivative with respect to the market price.

7.1.1 Introduction

In recent years, investigation of behavioral patterns of agents of mixed markets, in which state-owned (public, domestic, etc.) welfare-maximizing firms compete against profit-maximizing private (foreign) firms, has become more and more popular. For pioneering works on mixed oligopolies (see Merril and Schneider [225], Ruffin [276], Harris and Wiens [139], and Bös [24, 25]). Excellent surveys can be found in Vickers and Yarrow [308], De Frajas and Delbono [118], Nett [253].

S. Dempe et al., *Bilevel Programming Problems*, Energy Systems,
DOI 10.1007/978-3-662-45827-3_7

The interest in mixed oligopolies is high because of their importance to the economies of Europe (Germany, England and others), Canada and Japan (see Matsushima and Matsumura [224], for an analysis of "herd behavior" by private firms in many branches of the economy in Japan). There are examples of mixed oligopolies in the United States such as the packaging and overnight-delivery industries. Mixed oligopolies are also common in the East European and former Soviet Union transitional economies, in which competition among public and private firms existed or still exists in many industries such as banking, house loan, life insurance, airline, telecommunication, natural gas, electric power, automobile, steel, education, hospital, health care, broadcasting, railways and overnight-delivery. Moreover, according to Bös [25], Fershtman [107], Matshumura and Kanda [222], in many cases the government has held, or even still holds, a non-negligible proportion of shares in privatized firms, and there are firms with a mixture of private and public ownership. Since privatized firms with mixed ownership must respect the interests of private shareholders, they cannot be pure domestic social surplus maximizers. At the same time they must respect the interests of the government, so they cannot be pure profit-maximizers. By controlling the shares that it holds, the government may be able to indirectly control the activities of the privatized firm.

In the majority of the above-mentioned papers, the mixed oligopoly is studied in the framework of classical Cournot, Hotelling or Stackelberg models (cf. Matsushima and Matsumura [224], Matsumura [223], Cornes and Sepahvand [44]). It is well known (cf. for instance, Figuières et al. [110]) that the Nash equilibrium (including Cournot equilibrium as a particular case) is the outcome consistent with rational agents who take rival decisions as given when they optimize. Alternately, in the Stackelberg equilibrium there are two agents who take their decisions sequentially; the first agent to move is referred to as the leader, whereas the second mover is called the follower. The Stackelberg equilibrium is an outcome consistent with the follower's rational behavior given that he has observed the leader's move, and the leader's rational behavior who can infer what will be the follower's rational reaction to his current decision.

Conjectural variations equilibria (CVE) were introduced by Bowleyl [26] and Frisch [119, 120], as another possible solution concept in static games. According to this concept, agents behave as follows: each agent chooses his/her most favorable action taking into account that every rival's strategy is a conjectured function of her own strategy.

In the works by Bulavsky and Kalashnikov [37, 38, 152], a new scale of conjectural variations equilibria (CVE) was introduced and investigated, in which the conjectural variations (represented via the influence coefficients of each agent) affected the structure of the Nash equilibrium. In other words, we considered not only a classical Cournot competition but also a Cournot-type model with influence coefficient values different from 1 (as the influence coefficient 1 corresponds to the classical Cournot model). Various equilibrium existence and uniqueness results were obtained in the above-cited works.

For instance, in Isac et al. [152], the classical oligopoly model was extended to the conjectural oligopoly as follows. Instead of the classical Cournot assumptions, all producers $i = 1, 2, \ldots, n$, used the conjectural variations described below:

$$G_i(\eta) = G + (\eta - q_i)\, w_i\, (G, q_i)\,.$$

Here, G is the current total quantity of the product cleared in the market, q_i and η are, respectively, the present and the expected supplies by the ith agent, whereas $G_i(\eta)$ is the total cleared market volume *conjectured* by the ith agent as a response to changing his/her own supply from q_i to η. The conjecture function w_i was referred to as the ith agent's *influence quotient* (*coefficient*). Notice that the classical Cournot model assumes $w_i \equiv 1$ for all i. Under general enough assumptions concerning properties of the influence coefficients $w_i = w_i\,(G, q_i)$, cost functions $f_i = f_i(q_i)$, and the inverse demand (price) function $p = p(G)$, new existence and uniqueness results for the conjectural variations equilibrium (CVE) were obtained. This approach was further developed in Kalashnikov et al. [168, 175] with application to the mixed oligopoly model. Here again, all agents (both public and private companies) make their decisions based upon the model's data (inverse demand and cost functions) and their influence coefficients (conjectures) $w_i = w_i\,(G, q_i)$.

As is mentioned in Figuières et al. [110], Giocoli [130], the concept of conjectural variations has been the subject of numerous theoretical controversies (see e.g. Lindhi [204]). Nevertheless, economists have made extensive use of one form or the other of the CVE to predict the outcome of non-cooperative behavior in several fields of economics. The literature on conjectural variations has focused mainly on two-player games (cf. Figuières et al. [110]). The central concept of the theory is the notion of *conjecture*. The *variational conjecture* r_j usually describes player j's reaction, as anticipated by player i, to an *infinitesimal variation* of player i's strategy. This mechanism leads to the notion of a *conjectured reaction function* of the opponent. Given these conjectured reactions on part of the rivals, each agent optimizes his/her perceived payoff. This leads to the concept of a *conjectural best response function*. An equilibrium is obtained when no player has an interest in deviating from his/her strategy, i.e., his/her conjectural best response to the strategies of the other player.

The *consistency* (or, sometimes, "rationality") of the equilibrium is defined as the *coincidence* between the conjectural best response of each agent and the conjectured reaction function of the same. A conceptual difficulty arises when one considers consistency in the case of many agents (see, Figuières et al. [110]). The strongest notion of consistency requires that the conjectural best response of player i coincides with what the other players have conjectured about his/her reaction, that is, with one of their conjectured reaction functions. However, when n agents are present, there are n best response functions and $n(n-1)$ conjectures. Therefore, if $n > 2$, equilibrium is consistent only if all players have the same conjectures about player i's reaction. This is the approach followed explicitly by Başar and Olsder [7]; this assumption can be also found in Fershtman and Kamien [108] dealing with conjectures in differential games. In the literature on conjectural variations in static n-player games, the problem is usually implicitly addressed by assuming a complete identity of all the agents (cf. Laitner [198], Bresnahan [28] and references therein, Novshek [256]). Using a bit different approach, Perry [263] for oligopoly, Cornes and Sandler [43] and

Sugden [295] for public goods, consider a class of games where for each agent, the contributions of all other players to her payoff are aggregated. It is as if each agent plays against a unique (virtual) player representing the remaining agents.

To cope with this conceptual difficulty arising in many players models, Bulavsky [36] proposed a completely new approach. Instead of assuming the identity of the agents in the conjectural variation model of a homogeneous good market, it is supposed that each player makes conjectures not about the (optimal) response functions of the other players but only about the variations of the market price depending upon his infinitesimal output variations. Knowing the rivals' conjectures (called influence coefficients), each agent can realize certain verification procedure and check out if his influence coefficient is consistent with the others. Exactly the same verification formulas were obtained independently in Liu et al. [206] establishing the existence and uniqueness of consistent conjectural variation equilibrium in electricity market. However, they applied a much more difficult optimal control technique, searching only steady states as a final result (a similar technique was used in Driskill and McCafferty [95]. Moreover, they restricted the inverse demand function to a linear one, and the agents' cost functions to quadratic ones in their model, whereas the approach in Bulavsky [36] allows nonlinear and even non-differentiable demand functions and arbitrary (twice continuously differentiable) convex cost functions of the agents.

In this section, we extend the results obtained in Bulavsky [36] to a mixed oligopoly model. In the same manner as in Bulavsky and Kalashnikov [37, 38], we consider a conjectural variations oligopoly model, in which the degree of influence on the whole situation by each agent is modeled by special parameters (influence coefficients). However, in contrast to the models defined in Bulavsky and Kalashnikov [37, 38] and Kalashnikov et al. [168, 175], here, we follow the ideology of Bulavsky [35, 36] selecting the market clearing price p, rather than the producers' output, as an observable variable.

The section is organized as follows. In Sect. 7.1.2, we describe the mathematical model from Bulavsky [36] extended to the mixed oligopoly case and then, in Sect. 7.1.3, we define the concept of exterior equilibrium, i.e., a conjectural variations equilibrium (CVE) with the influence coefficients fixed in an exogenous form. The existence and uniqueness theorem for this kind of CVE ends the subsection. Section 7.1.4 deals with the more advanced concept of interior equilibrium, which is defined as the exterior equilibrium with consistent conjectures (influence coefficients). The consistency criterion, the consistency verification procedure, and the existence theorem for the interior equilibrium are formulated in the same Sect. 7.1.4. To provide the tools for the future research concerning the interrelationships between the demand structure (with not necessarily smooth demand function) and the CVEs with consistent conjectures (influence coefficients), the behavior of the latter as functions of certain parameter (governed by the derivative by p of the demand function $G = G(p)$) is studied in Theorem 7.3 completing Sect. 7.1.4. Finally, Sect. 7.1.5 contains the results of numerical experiments with a test model of an electricity market from Liu et al. [206], with and without a public company among the agents.

7.1.2 Model Specification

Consider a market of a homogeneous good (natural gas, oil, electricity, timber, etc.) with no less than 3 producers/suppliers with cost functions $f_i = f_i(q_i)$, $i = 0, 1, \dots, n$, where $n \geq 2$, and q_i is the output/supply brought by producer i, $i = 0, 1, \dots, n$. Consumers' demand is described by a demand function $G = G(p)$, whose argument p is the market clearing price. An active demand value D is nonnegative and does not depend upon the price. We will reflect the equilibrium between the demand and supply for a given (clearing) price p by the following balance equality

$$\sum_{i=0}^{n} q_i = G(p) + D. \tag{7.1}$$

We assume the following properties of the model's data.

A1. The demand function $G = G(p) \geq 0$ defined for the (clearing) price values $p \in (0, +\infty)$ is non-increasing and continuously differentiable. □

A2. For each producer/supplier $i = 0, 1, \dots, n$, its cost function $f_i = f_i(q_i)$ is quadratic, i.e.,

$$f_i(q_i) = \frac{1}{2} a_i q_i^2 + b_i q_i, \tag{7.2}$$

with $a_i > 0$, $b_i > 0$, $i = 0, 1, \dots, n$. Moreover, we assume that

$$b_0 \leq \max_{1 \leq i \leq n} b_i. \tag{7.3}$$

Each private (or, foreign) producer i, $i = 1, \dots, n$, chooses his/her output volume $q_i \geq 0$ so as to maximize his/her net profit function $\pi(p, q_i) := p \cdot q_i - f_i(q_i)$. On the other hand, the public (or, domestic) company number $i = 0$ selects its production value $q_0 \geq 0$ so as to maximize domestic social surplus defined as the difference between the consumer surplus, the private (foreign) companies' total revenue, and the public (domestic) firm's production costs:

$$S(p; q_0, q_1, \dots, q_n) = \int_0^{\sum_{i=0}^{n} q_i} p(x)dx - p \cdot \left(\sum_{i=1}^{n} q_i \right) - b_0 q_0 - \frac{1}{2} a_0 q_0^2. \tag{7.4}$$

Now we postulate that the agents (both public and private) assume that their variation of production volumes may affect the price value p. The latter assumption could be implemented by accepting a conjectured dependence of fluctuations of the price p upon the variations of the (individual) output values q_i. Having that done, the first order maximum condition to describe the equilibrium would have the form: For the public company (with $i = 0$)

$$\frac{\partial S}{\partial q_0} = p - \left(\sum_{i=1}^{n}\right)\frac{\partial p}{\partial q_0} - f_0'(q_0)\begin{cases} = 0, & \text{if } q_0 > 0; \\ \leq 0, & \text{if } q_0 = 0; \end{cases} \tag{7.5}$$

and

$$\frac{\partial \pi_i}{\partial q_i} = p + q_i\frac{\partial p}{\partial q_i} - f_i'(q_i)\begin{cases} = 0, & \text{if } q_i > 0; \\ \leq 0, & \text{if } q_i = 0, \end{cases} \quad \text{for } i = 1, \ldots, n. \tag{7.6}$$

Therefore, we see that to describe the behavior of agent i and treat the maximum (equilibrium) conditions, it is enough to trace the derivative $\partial p/\partial q_i = -v_i$ rather than the full dependence of p upon q_i. (We introduce the minus here in order to deal with nononegative values of v_i, $i = 0, 1, \ldots, n$.) Of course, the conjectured dependence of p on q_i must provide (at least local) concavity of the ith agent's conjectured profit as a function of its output. Otherwise, one cannot guarantee the profit to be maximized (but not minimized). As we suppose that the cost functions $f_i = f_i(q_i)$ are quadratic and convex, then, for $i = 1, \ldots, n$, the concavity of the product $p \cdot q_i$ with respect to the variation η_i of the current production volume will do. For instance, it is sufficient to assume the coefficient v_i (from now on referred to as the ith agent's *influence coefficient*) to be nonnegative and constant. Then the conjectured local dependence of the agent's net profit upon the production output η_i has the form $[p - v_i(\eta_i - q_i)]\eta_i - f_i(\eta_i)$, while the maximum condition at $\eta_i = q_i$ is provided by the relationships

$$\begin{cases} p = v_i q_i + b_i + a_i q_i, & \text{if } q_i > 0; \\ p \leq b_i, & \text{if } q_i = 0. \end{cases} \tag{7.7}$$

Similarly, the public company conjectures the local dependence of domestic social surplus on its production output η_0 in the form

$$\int_0^{\eta_0 + \sum_{i=1}^{n} q_i} p(x)dx - [p - v_0(\eta_0 - q_0)] \cdot \left(\sum_{i=1}^{n} q_i\right) - b_0 - a_0 q_0, \tag{7.8}$$

which allows one to write down the (domestic social surplus) maximum condition at $\eta_0 = q_0$ as follows:

$$\begin{cases} p = -v_0 \sum_{i=1}^{n} q_i + b_0 + a_0 q_0, & \text{if } q_0 > 0; \\ p \leq -v_0 \sum_{i=1}^{n} q_i + b_0, & \text{if } q_0 = 0. \end{cases} \tag{7.9}$$

Were the agents' conjectures given exogenously (like it was assumed in Bulavsky and Kalashnikov [37, 38]), we would allow all the influence coefficients v_i to be functions of q_i and p. However, we use the approach from the papers Bulavski [35, 36], where the (justified, or consistent) conjectures are determined simultaneously with the equilibrium price p and output values q_i by a special verification procedure. In the latter case, the influence coefficients are the scalar parameters determined only at the equilibrium. In what follows, such equilibrium is referred to as *interior* one and is described by the set of variables and parameters $(p, q_0, q_1, \ldots, q_n, v_0, v_1, \ldots, v_n)$.

7.1.3 Exterior Equilibrium

Before we introduce the verification procedure, we need an initial notion of equilibrium called *exterior* (cf. Bulavski [36]) with the parameters (influence coefficients) $v_i, i = 0, 1, \ldots, n$ given exogenously.

Definition 7.1 The collection $(p, q_0, q_1, \ldots, q_n)$ is called *exterior* equilibrium for given influence coefficients $(v_0, v_1, \ldots, v_n)$, if the market is balanced, i.e., condition (7.1) is satisfied, and for each i, $i = 0, 1, \ldots, n$, the maximum conditions (7.7) and (7.9) are valid. □

In what follows, we are going to consider only the case when the list of really producing/supplying participants is fixed (i.e., it does not depend upon the values of the model's parameters). In order to guarantee this property, we make the following additional assumption.

A3. For the price value $p_0 := \max_{1\le j\le n} b_j$, the following (strict) inequality holds:

$$\sum_{i=0}^{n} \frac{p_0 - b_i}{a_i} < G(p_0). \tag{7.10}$$

Remark 7.1 The latter assumption, together with assumptions **A1** and **A2**, guarantees that for all nonnegative values of v_i, $i = 1, \ldots, n$, and for $v_0 \in [0, \overline{v}_0)$, where

$$0 < \overline{v}_0 = \begin{cases} a_0 \left[\dfrac{G(p_0) - \dfrac{p_0 - b_0}{a_0}}{\sum_{i=1}^{n} \dfrac{p_i - b_i}{a_i}} - 1 \right], & \text{if } \displaystyle\sum_{i=1}^{n} \frac{p_i - b_i}{a_i} > 0; \\ +\infty, & \text{otherwise}, \end{cases} \tag{7.11}$$

there always exists a unique solution of the optimality conditions (7.7) and (7.9) satisfying the balance equality (7.1), i.e., the exterior equilibrium. Moreover, conditions (7.1), (7.7) and (7.9) can hold simultaneously if, and only if $p > p_0$, that is, if and only if all outputs q_i are strictly positive, $i = 0, 1, \ldots, n$. Indeed, if $p > p_0$ then it is evident that neither inequalities $p \le b_i$, $i = 1, \ldots, n$, from (7.7), nor $p \le -v_0 \sum_{i=1}^{n} q_i + b_0$ from ((7.9) are possible, which means that none of q_i, $i = 0, 1, \ldots, n$, satisfying (7.7) and (7.9) can be zero.

Conversely, if all q_i, satisfying (7.7) and (7.9) are positive ($q_i > 0, i = 0, 1, \ldots, n$), then it is straightforward from conditions (7.7) that

$$p = v_i q_i + b_i + a_i q_i > b_i, \quad i = 1, \ldots, n;$$

hence $p > \max_{1 \le i \le n} = p_0$. □

The following theorem is the main result of this subsection and a tool for the introduction of the concept of interior equilibrium in the next subsection.

Theorem 7.1 *Under assumptions* **A1**, **A2** *and* **A3**, *for any* $D \ge 0$, $v_i \ge 0$, $i = 1, \ldots, n$, *and* $v_0 \in [0, v_0)$, *there exists a unique exterior equilibrium state* $(p, q_0, q_1, \ldots, q_n)$, *which depends continuously on the parameters* $(D, v_0, v_1, \ldots, v_n)$. *The equilibrium price* $p = p(D, v_0, v_1, \ldots, v_n)$ *as a function of these parameters is differentiable with respect to both D and* v_i, $i = 0, 1, \ldots, n$. *Moreover,* $p = p(D, v_0, v_1, \ldots, v_n) > p_0$, *and*

$$\frac{\partial p}{\partial D} = \frac{1}{\dfrac{v_0 + a_0}{a_0} \sum_{i=0}^{n} \dfrac{1}{v_i + a_i} - G'(p)}. \tag{7.12}$$

Proof Due to assumptions **A1–A3**, for any fixed collection of conjectures $v = (v_0, v_1, \ldots, v_n) \ge 0$, the equalities in the optimality conditions (7.7) and (7.9) determine the optimal response (to the existing clearing price) values of the producers/suppliers as continuously differentiable (with respect to p) functions $q_i = q_i(p; v_0, \ldots, v_n)$ defined over the interval $p \in [p_0, +\infty)$ by the following explicit formulas:

$$q_0 := \frac{p - b_0}{a_0} + \frac{v_0}{a_0} \sum_{i=1}^{n} \frac{p - b_i}{v_i + a_i}, \tag{7.13}$$

and

$$q_i := \frac{p - b_i}{v_i + a_i}, \quad i = 1, \ldots, n. \tag{7.14}$$

Moreover, the partial derivatives of the optimal response functions are positive:

$$\frac{\partial q_0}{\partial p} = \frac{1}{a_0} + \frac{v_0}{a_0} \sum_{i=1}^{n} \frac{1}{v_i + a_i} \ge \frac{1}{a_0} > 0, \tag{7.15}$$

and

$$\frac{\partial q_i}{\partial p} = \frac{1}{v_i + a_i} > 0, \quad i = 1, \ldots, n. \tag{7.16}$$

Therefore, the total production volume function

$$Q\left(p; v_0, v_1, \ldots, v_n\right) = \sum_{i=0}^{n} q_i\left(p; v_0, v_1, \ldots, v_n\right)$$

is continuous and strictly increasing by p. According to assumption **A3**, this function's value at the point $p = p_0$ is strictly less that $G(p_0)$. Indeed, from (7.13) and (7.14) we have:

(A) If $\sum_{i=1}^{n} \frac{p_0 - b_i}{v_i + a_i} > 0$, then

$$\begin{aligned} Q\left(p_0; v_0, v_1, \ldots, v_n\right) &= \sum_{i=0}^{n} q_i\left(p_0; v_0, v_1, \ldots, v_n\right) \\ &= \frac{p_0 - b_0}{a_0} + \frac{v_0}{a_0} \sum_{i=1}^{n} \frac{p_0 - b_i}{v_i + a_i} + \sum_{i=1}^{n} \frac{p_0 - b_i}{v_i + a_i} \\ &= \frac{p_0 - b_0}{a_0} + \frac{v_0 + a_0}{a_0} \sum_{i=1}^{n} \frac{p_0 - b_i}{v_i + a_i} \\ &< \frac{p_0 - b_0}{a_0} + \frac{\overline{v}_0 + a_0}{a_0} \sum_{i=1}^{n} \frac{p_0 - b_i}{a_i} \\ &= \frac{p_0 - b_0}{a_0} + \left[\frac{G(p_0) - \frac{p_0 - b_0}{a_0}}{\sum_{i=1}^{n} \frac{p_0 - b_i}{a_i}} \right] \sum_{i=1}^{n} \frac{p_0 - b_i}{a_i} \\ &= \frac{p_0 - b_0}{a_0} + G(p_0) - \frac{p_0 - b_0}{a_0} = G(p_0). \end{aligned}$$

(B) Otherwise, i.e., if $\sum_{i=1}^{n} \frac{p_0 - b_i}{v_i + a_i} = 0$, one has:

$$\begin{aligned} Q\left(p_0; v_0, v_1, \ldots, v_n\right) &= \sum_{i=0}^{n} q_i\left(p_0; v_0, v_1, \ldots, v_n\right) \\ &= \frac{p_0 - b_0}{a_0} + \frac{v_0}{a_0} \sum_{i=1}^{n} \frac{p_0 - b_i}{v_i + a_i} + \sum_{i=1}^{n} \frac{p_0 - b_i}{v_i + a_i} \\ &= \frac{p_0 - b_0}{a_0} < G(p_0) \end{aligned}$$

for any $v_i \geq 0$, $i = 1, \ldots, n$, and $v_0 \in [0, \overline{v}_0)$.

On the other hand, the total output supply $Q = Q(p; v_0, v_1, \ldots, v_n)$ clearly tends to $+\infty$ when $p \to +\infty$. Now define

$$p_* := \sup\{p : \; Q(p; v_0, v_1, \ldots, v_n) \leq G(p) + D\}. \tag{7.17}$$

Since both functions $Q(p; v_0, v_1, \ldots, v_n)$ and $G(p) + D$ are continuous with respect to p, the former increases unboundedly and the latter, vice versa, is non-increasing by p over the whole ray $[p_0, +\infty)$, then, first, the value of p_* is finite ($p_* < +\infty$), and second, by definition (7.17) and the continuity of both functions,

$$Q(p_*; v_0, v_1, \ldots, v_n) \leq G(p_*) + D.$$

Now we demonstrate that the strict inequality $Q(p_*; v_0, v_1, \ldots, v_n) < G(p_*) + D$ cannot happen. Indeed, suppose on the contrary that the latter strict inequality holds. Then the continuity of the involved functions implies that for some values $p > p_*$ sufficiently close to p_*, the same relationship is true: $Q(p; v_0, v_1, \ldots, v_n) < G(p) + D$, which contradicts definition (7.17). Therefore, the exact equality holds

$$Q(p_*; v_0, v_1, \ldots, v_n) = G(p_*) + D, \tag{7.18}$$

which, in its turn, means that the values p_* and $q_i^* = q_i(p_*; v_0, v_1, \ldots, v_n)$, $i = 0, \ldots, n$, determined by formulas (7.13) and (7.14) form an exterior equilibrium state for the collection of influence coefficients $v = (v_0, v_1, \ldots, v_n)$. The uniqueness of this equilibrium follows from the fact that the function $Q = Q(p; v_0, v_1, \ldots, v_n)$ strictly increases while the demand function $G(p) + D$ is non-increasing with respect to p. Indeed, these facts combined with (7.18) yield that $Q(p; v_0, v_1, \ldots, v_n) < G(p) + D$ for all $p \in (p_0, p_*)$, whereas $Q(p; v_0, v_1, \ldots, v_n) > G(p) + D$ when $p > p_*$. To conclude, the equilibrium price p_* and hence, the equilibrium outputs $q_i^* = q_i(p_*; v_0, v_1, \ldots, v_n)$, $i = 0, \ldots, n$, calculated by formulas (7.13) and (7.14), are determined uniquely.

Now we establish the continuous dependence of the equilibrium price (and hence, the equilibrium output volumes, too) upon the parameters $(D, v_0, \ldots, v_n)$. To do that, we substitute expressions (7.13) and (7.14) for $q_i = (p; v_0, \ldots, v_n)$ into the balance equality (7.1) and come to the following relationship:

$$\begin{aligned} q_0 + \sum_{i=1}^{n} q_i - G(p) - D &= \left(\frac{p - b_0}{a_0} + \frac{v_0}{a_0} \sum_{i=1}^{n} \frac{p - b_i}{v_i + a_i} \right) \\ &\quad + \sum_{i=1}^{n} \frac{p - b_i}{v_i + a_i} - G(p) - D \\ &= p\left(\frac{1}{a_0} + \frac{v_0 + a_0}{a_0} \sum_{i=1}^{n} \frac{1}{v_i + a_i} \right) - \frac{b_0}{a_0} \\ &\quad - \frac{v_0 + a_0}{a_0} \sum_{i=1}^{n} \frac{b_i}{v_i + a_i} - G(p) - D = 0. \end{aligned} \tag{7.19}$$

Introduce the function

$$\Gamma\,(p; v_0, v_1, \ldots, v_n, D) = p\left(\frac{1}{a_0} + \frac{v_0 + a_0}{a_0}\sum_{i=1}^{n}\frac{1}{v_i + a_i}\right) - \frac{b_0}{a_0} - \frac{v_0 + a_0}{a_0}\sum_{i=1}^{n}\frac{b_i}{v_i + a_i} - G(p) - D$$

and rewrite the last equality in (7.19) as the functional equation

$$\Gamma\,(p; v_0, v_1, \ldots, v_n, D) = 0. \tag{7.20}$$

As the partial derivative of the latter function with respect to p is (always) positive:

$$\frac{\partial \Gamma}{\partial p} = \frac{1}{a_0} + \frac{v_0 + a_0}{a_0}\sum_{i=1}^{n}\frac{1}{v_i + a_i} - G'(p) \geq \frac{1}{a_0} > 0,$$

one can apply Implicit Function Theorem and conclude that the equilibrium (clearing) price p treated as an explicit function $p = p\,(v_0, v_1, \ldots, v_n, D)$ is continuous and, in addition, differentiable with respect to all the parameters $v_0, v_1, \ldots, v_n, D$. Moreover, the partial derivative of the equilibrium price p with respect to D can be calculated from the full derivative equality

$$\frac{\partial \Gamma}{\partial p} \cdot \frac{\partial p}{\partial D} + \frac{\partial \Gamma}{\partial D} = 0,$$

finally yielding the desired formula (7.12)

$$\frac{\partial p}{\partial D} = \frac{1}{\dfrac{v_0 + a_0}{a_0}\sum_{i=0}^{n}\dfrac{1}{v_i + a_i} - G'(p)},$$

and thus completing the proof. □

7.1.4 *Interior Equilibrium*

Now we are ready to define the concept of interior equilibrium. To do that, we first describe the procedure of verification of the influence coefficients v_i as it was given in Bulavski [36]. Assume that we have an exterior equilibrium state $(p, q_0, q_1, \ldots, q_n)$ that occurs for some feasible $v = (v_0, v_1, \ldots, v_n)$ and $D \geq 0$. One of the producers, say k, $0 \leq k \leq n$, temporarily changes his/her behavior by *abstaining* from maximization of the conjectured profit (or domestic social surplus, as is in case $k = 0$) and making small fluctuations (variations) around his/her equilibrium output volume q_k. In mathematical terms, the latter is tantamount to restricting the model agents to the subset $I_{-k} := \{0 \le i \le n : i \ne k\}$ with the active demand reduced to $D_{-k} := D - q_k$.

A variation δq_k of the production output by agent k is then equivalent to the active demand variation in form $\delta D_{-k} := -\delta q_k$. If we consider these variations being infinitesimal, we assume that by observing the corresponding variations of the equilibrium price, agent k can evaluate the derivative of the equilibrium price with respect to the active demand in the reduced market, which clearly coincides with his/her influence coefficient.

When applying formula (7.12) from Theorem 8.1 to evaluate the player k conjecture (influence coefficient) v_k, one has to remember that agent k is temporarily absent from the equilibrium model, hence one has to *exclude* from all the sums the term with number $i = k$. Keeping that in mind, we come to the following *consistency criterion*.

7.1.4.1 Consistency Criterion

At an exterior equilibrium $(p, q_0, q_1, \dots, q_n)$, the influence coefficients v_k, $k = 0, 1, \dots, n$, are referred to as *consistent* if the following equalities hold:

$$v_0 = \frac{1}{\sum_{i=1}^{n} \frac{1}{v_i + a_i} - G'(p)}, \tag{7.21}$$

and

$$v_i = \frac{1}{\frac{v_0 + a_0}{a_0} \sum_{j=0, j \ne i}^{n} \frac{1}{v_j + a_j} - G'(p)}, \quad i = 1, \dots, n. \tag{7.22}$$

Now we are in a position to define the concept of an interior equilibrium.

Definition 7.2 The collection $(p, q_0, \dots, q_n, v_0, \dots, v_n)$ where $v_i \ge 0$, $i = 0, 1, \dots, n$, is referred to as the *interior equilibrium* if, for the coefficients $(v_0, v_1, \dots, v_n)$ the collection $(p, q_0, \dots, q_n)$ is an exterior equilibrium state, and the consistency criterion is satisfied for all $k = 0, 1, \dots, n$. □

Remark 7.2 If all the agents are profit-maximizing private companies, then formulas (7.21)–(7.22) reduce to the uniform ones obtained independently in Bulavski [36] and Lui et al. [206]:

$$v_i = \frac{1}{\sum_{j \in I \setminus \{i\}} \frac{1}{v_j + a_j} - G'(p)}, \quad i \in I, \tag{7.23}$$

where I is an arbitrary (finite) list of the participants of the model. □

The following theorem is an extension of Theorem 2 in Bulavski [36] to the case of a mixed oligopoly.

Theorem 7.2 *Under assumptions* **A1**, **A2**, *and* **A3**, *there exists the interior equilibrium.*

Proof We are going to show that there exist $v_0 \in [0, \bar{v}_0)$; $v_i \geq 0$, $i = 1, \ldots, n$; $q_i \geq 0$, $i = 0, 1, \ldots, n$, and $p > p_0$ such that the vector $(p; q_0, \ldots, q_n; v_0, \ldots, v_n)$ provides for the interior equilibrium. In other words, the vector $(p, q_0, \ldots, q_n)$ is an exterior equilibrium state, and in addition, equalities (7.21)–(7.22) hold. For a technical purpose, let us introduce a parameter α so that $G'(p) := \frac{\alpha}{1+\alpha}$ for appropriate values of $\alpha \in [-1, 0]$, and then rewrite the right-hand sides of formulas (7.21)–(7.22) in the following (equivalent) form:

$$F_0\left(\alpha; v_0, \ldots, v_n\right) := \frac{1+\alpha}{(1+\alpha)\sum_{i=1}^{n} \dfrac{1}{v_i + a_i} - \alpha}, \tag{7.24}$$

and

$$F_i\left(\alpha; v_0, \ldots, v_n\right) := \frac{1+\alpha}{(1+\alpha)\dfrac{v_0 + a_0}{a_0}\sum_{j=0, j\neq i}^{n} \dfrac{1}{v_j + a_j} - \alpha}, \quad i = 1, \ldots, n. \tag{7.25}$$

Since $v_i \geq 0$, $a_i > 0$, $i = 0, 1, \ldots, n$, and $\alpha \in [-1, 0]$, the functions F_i, $i = 0, 1, \ldots, n$, are well-defined and continuous with respect to their arguments over the corresponding domains. Now let us introduce an auxiliary function $\Phi : [-1, 0] \times R_+^{n+1}$ as follows. For arbitrary $\alpha \in [-1, 0]$ and $(v_0, v_1, \ldots, v_n) \in [0, \bar{v}_0) \times R_+^n$, find the (uniquely determined, according to Theorem 8.1) exterior equilibrium vector $(p, q_0, q_1, \ldots, q_n)$ and calculate the derivative $G'(p)$ at the equilibrium point p. Then define the value of the function Φ as below:

$$\Phi\left(\alpha; v_0, v_1, \ldots, v_n\right) := \hat{\alpha} = \frac{G'(p)}{1 - G'(p)} \in [-1, 0]. \tag{7.26}$$

When introducing this auxiliary function Φ, we do not indicate explicitly its dependence upon D, because we are not going to vary D while proving the theorem. As the derivative $G'(p)$ is continuous by p (see assumption **A1**), and the equilibrium price $p = p\left(v_0, v_1, \ldots, v_n\right)$, in its turn, is a continuous function (cf. Theorem 8.1), then the function Φ is continuous as a superposition of continuous functions. (We also notice that its dependence upon α is fictitious.) To finish the proof, let us compose a mapping $H := (\Phi, F_0, F_1, \ldots, F_n) : [-1, 0] \times R_+^{n+1} \to [-1, 0] \times R_+^{n+1}$ and select a convex compact that is mapped by H into itself. The compact is constructed as follows: First, set $s := \max\{\bar{v}_0, a_0, a_1, \ldots, a_n\}$. Second, formulas (7.24)–(7.25) yield the following relationships: If $\alpha = -1$, then

$$F_0(-1, v_0, v_1, \ldots, v_n) = 0, \tag{7.27}$$

$$F_i(-1, v_0, v_1, \ldots, v_n) = 0, \quad i = 1, \ldots, n, \tag{7.28}$$

whereas for $\alpha \in (-1, 0]$ one has

$$\begin{aligned} 0 \le F_0(\alpha, v_0, v_1, \ldots, v_n) &= \frac{1+\alpha}{(1+\alpha)\sum_{i=1}^{n} \frac{1}{v_i + a_i} - \alpha} \\ &\le \frac{1+\alpha}{(1+\alpha)\sum_{i=1}^{n} \frac{1}{v_i + a_i}} = \frac{1}{\sum_{i=1}^{n} \frac{1}{v_i + a_i}} \le \frac{1}{\sum_{i=1}^{n} \frac{1}{v_i + s}}; \end{aligned} \tag{7.29}$$

and

$$\begin{aligned} 0 \le F_i(\alpha, v_0, v_1, \ldots, v_n) &= \frac{1+\alpha}{(1+\alpha)\frac{v_0 + a_0}{a_0}\sum_{j=0, j\ne i}^{n} \frac{1}{v_j + a_j} - \alpha} \\ &\le \frac{1}{\frac{v_0 + a_0}{a_0}\sum_{j=0, j\ne i}^{n} \frac{1}{v_j + a_j}} \le \frac{1}{\sum_{j=0, j\ne i}^{n} \frac{1}{v_j + a_j}} \\ &\le \frac{1}{\sum_{j=0, j\ne i}^{n} \frac{1}{v_j + s}}, \quad i = 1, \ldots, n. \end{aligned} \tag{7.30}$$

Relationships (7.27)–(7.30) clearly imply that for any $\alpha \in [-1, 0]$, if $0 \le v_j \le \frac{s}{n-1}$, $j = 0, 1, \ldots, n$, then the values of $F_j(\alpha, v_0, \ldots, v_n)$, $j = 0, \ldots, n$, drop within the same (closed) interval $\left[0, \frac{s}{n-1}\right]$. Therefore, we have just established that the continuous mapping $H := (\Phi, F_0, F_1, \ldots, F_n)$ maps the compact convex subset $\Omega := [-1, 0] \times \left[0, \frac{s}{n-1}\right]^{n+1}$ into itself. Thus, by Brouwer Fixed Point Theorem, the mapping H has a fixed point $(\alpha, v_0, \ldots, v_n)$, that is,

$$\begin{cases} \Phi(\alpha, v_0, v_1, \ldots, v_n) = \alpha, \\ F_0(\alpha, v_0, v_1, \ldots, v_n) = v_0, \\ F_1(\alpha, v_0, v_1, \ldots, v_n) = v_1, \\ \vdots \\ F_n(\alpha, v_0, v_1, \ldots, v_n) = v_n. \end{cases} \tag{7.31}$$

Now, for the thus obtained influence coefficients $v = (v_0, v_1, \ldots, v_n) \in [0, \overline{v}_0) \times R_+^n$, there exists (uniquely, by Theorem 8.1) the exterior equilibrium $(p, q_0, q_1, \ldots, q_n)$. Hence, we can immediately conclude (from (8.51) and the definition of function Φ) that $G'(p) = \frac{\alpha}{1+\alpha}$, and therefore, the influence coefficients satisfy conditions (7.21)–(7.22). So, according to Definition 7.2, the just constructed vector $(p; q_0, \ldots, q_n; v_0, \ldots, v_n)$ is the desired interior equilibrium. The proof is complete. □

7.1.4.2 Properties of Influence Coefficients

In our future research, we are going to extend the obtained results to the case of non-differentiable demand functions. However, some of the necessary technique can be developed now, in the differentiable case. To do that, we denote the value of the demand function's derivative by $\tau := G'(p)$ and rewrite the consistency Eqs. (7.21)–(7.22) in the following form:

$$v_0 = \frac{1}{\sum_{i=1}^{n} \frac{1}{v_i + a_i} - \tau}, \tag{7.32}$$

and

$$v_i = \frac{1}{\frac{v_0 + a_0}{a_0} \sum_{j=0, j \neq i}^{n} \frac{1}{v_j + a_j} - \tau}, \quad i = 1, \ldots, n, \tag{7.33}$$

where $\tau \in (-\infty, 0]$. When $\tau \to -\infty$ then system (7.32)–(7.33) has the unique limit solution $v_j = 0,\ j = 0, 1, \ldots, n$. For all the finite values of τ, we establish the following result.

Theorem 7.3 *For any $\tau \in (-\infty, 0]$, there exists a unique solution of equations (7.32)–(7.33) denoted by $v_k = v_k(\tau)$, $k = 0, 1, \ldots, n$, continuously depending upon τ. Furthermore, $v_k(\tau) \to 0$ when $\tau \to -\infty$, $k = 0, \ldots, n$, and $v_0 = v_0(\tau)$ strictly increases until $v_0(0)$ if τ grows up to zero.*

Proof Similar to the proof of Theorem 7.2, we introduce the auxiliary functions

$$F_0(\tau; v_0, \ldots, v_n) := \frac{1}{\sum_{i=1}^{n} \frac{1}{v_i + a_i} - \tau} = v_0, \tag{7.34}$$

and

$$F_i(\tau; v_0, \ldots, v_n) := \frac{1}{\frac{v_0 + a_0}{a_0} \sum_{j=0, j \neq i}^{n} \frac{1}{v_j + a_j} - \tau} = v_i, \quad i = 1, \ldots, n, \tag{7.35}$$

and set $s := \max\{a_0, a_1, \ldots, a_n\}$. It is easy to check that for any fixed value of $\tau \in (-\infty, 0]$, the vector-function $\mathrm{d} := (F_0, F_1, \ldots, F_n)$ maps the n-dimensional cube $M := \left[0, \frac{s}{n-1}\right]^n$ into itself. Now we show that subsystem (7.35) has a unique solution $v(v_0, \tau) = (v_1(v_0, \tau), \ldots, v_n(v_0, \tau))$ for any fixed $\tau \in (-\infty, 0]$ and $v_0 \geq 0$; moreover, the vector-function $v = v(v_0, \tau)$ is continuously differentiable with respect to both variables v_0 and τ. The Jacobi matrix of the mapping $\mathrm{d} = (F_0, F_1, \ldots, F_n)$, that is, the matrix $J := \left(\frac{\partial F_i}{\partial v_j}\right)_{i=1, j=1}^{n, n}$ has the following entries:

$$\frac{\partial F_i}{\partial v_j} = \begin{cases} 0, & \text{for } j = i; \\ \frac{v_0 + a_0}{a_0} \cdot \frac{F_i^2}{\left(v_j + a_j\right)^2}, & \text{for } j \neq i. \end{cases} \tag{7.36}$$

Thus, matrix J is nonnegative and non-decomposable. Now let us estimate the sums of the matrix entries in each row $i = 1, 2, \ldots, n$:

$$\sum_{k=1}^{n} \frac{\partial F_i}{\partial v_k} = \frac{v_0 + a_0}{a_0} F_i^2 \cdot \sum_{k=1, k\neq i}^{n} \frac{1}{(v_k + a_k)^2} \leq \frac{\frac{v_0 + a_0}{a_0} \sum_{k=1, k\neq i}^{n} \frac{1}{(v_k + a_k)^2}}{\left(\frac{v_0 + a_0}{a_0} \sum_{k=1, k\neq i}^{n} \frac{1}{v_k + a_k}\right)^2}$$

$$= \frac{a_0}{v_0 + a_0} \cdot \frac{\sum_{k=1, k\neq i}^{n} \frac{1}{(v_k + a_k)^2}}{\left(\sum_{k=1, k\neq i}^{n} \frac{1}{v_k + a_k}\right)^2} = R_i(v_1, \ldots, v_n; v_0) < 1. \tag{7.37}$$

For any fixed value $v_0 \geq 0$ (treated as a parameter), the above-mentioned functions $R_i(v_1, \ldots, v_n; v_0)$, $i = 1, \ldots, n$, are defined on the cube M, continuously depend upon the variables $v_1, \ldots, v_n$, and take only positive values strictly less than 1. Therefore, their maximum values achieved on the compact cube M are also strictly lower than 1, which implies that the matrix $(I - J)$ is invertible (here, I is the n-dimensional unit matrix), and the mapping $\mathrm{d} := (F_1, \ldots, F_n)$ defined on M is a strictly contracting mapping in the cubic norm (i.e., $\|\cdot\|_\infty$-norm). The latter allows to conclude that for any fixed values of $\tau \in (-\infty, 0]$ and $v_0 \geq 0$, the equation subsystem (7.35) has a unique solution $v(v_0, \tau) = (v_1(v_0, \tau), \ldots, v_n(v_0, \tau))$. Since $\det(I - J) \neq 0$ for any $\tau \in (-\infty, 0]$, Implicit Function Theorem also guarantees that $v(v_0, \tau)$ is continuously differentiable by both variables.

In order to establish the monotone increasing dependence of the solution $v(v_0, \tau)$ of subsystem (7.35) upon τ for any fixed value $v_0 \geq 0$, let us differentiate (7.35) with respect to τ to yield

$$\frac{\partial v_i}{\partial \tau} = F_i^2 \left[1 + \frac{v_0 + a_0}{a_0} \sum_{k=1, k \neq i}^{n} \frac{\frac{\partial v_k}{\partial \tau}}{(v_k + a_k)^2} \right], \quad i = 1, \ldots, n. \tag{7.38}$$

Rewriting (7.38) in the vector form, we come to

$$v'_\tau = J v'_\tau + F^2, \tag{7.39}$$

where

$$v'_\tau := \left(\frac{\partial v_1}{\partial \tau}, \ldots, \frac{\partial v_n}{\partial \tau} \right)^T \quad \text{and} \quad F^2 := \left(F_1^2, \ldots, F_n^2 \right)^T > 0. \tag{7.40}$$

Since all entries of the inverse of matrix $(I - J)$ are nonnegative (the latter is due to the matrix $(I - J)$ being an M-matrix, cf. e.g. Berman and Plemmons [21]) and the inverse matrix $(I - J)^{-1}$ has no zero rows, then (7.39)–(7.40) imply

$$v'_\tau = (I - J)^{-1} F^2 > 0, \tag{7.41}$$

that is, each component of the solution vector $v\,(v_0, \tau)$ of subsystem (7.35) is a strictly increasing function of τ for each fixed value of $v_0 \geq 0$. Moreover, the straightforward estimates

$$v_i\,(v_0, \tau) \leq -\frac{1}{\tau}, \quad i = 1, \ldots, n, \tag{7.42}$$

bring about the limit relationships shown below:

$$v_i\,(v_0, \tau) \to 0 \text{ as } \tau \to -\infty, \quad i = 1, \ldots, n, \quad \text{for any fixed } v_0 \geq 0. \tag{7.43}$$

To order to establish the monotone (decrease) dependence of the solution $v\,(v_0, \tau)$ of subsystem (7.35) upon $v_0 \geq 0$ for each fixed value of $\tau \in (-\infty, 0]$, we differentiate (7.35) with respect to v_0 to get:

$$\begin{aligned} \frac{\partial v_i}{\partial v_0} &= -F_i^2 \left[\frac{1}{a_0} \sum_{k=1, k \neq i}^{n} \frac{1}{v_k + a_k} - \frac{v_0 + a_0}{a_0} \sum_{k=1, k \neq i}^{n} \frac{\frac{\partial v_k}{\partial v_0}}{(v_k + a_k)^2} \right] \\ &= F_i^2 \frac{v_0 + a_0}{a_0} \sum_{k=1, k \neq i}^{n} \frac{\frac{\partial v_k}{\partial v_0}}{(v_k + a_k)^2} - \frac{F_i^2}{a_0} \sum_{k=1, k \neq i}^{n} \frac{1}{v_k + a_k}, \quad i = 1, \ldots, n. \end{aligned} \tag{7.44}$$

Again, rearrange these equalities into a system of equations

$$v'_{v_0} = J v'_{v_0} - Q, \tag{7.45}$$

where

$$v'_{v_0} := \left(\frac{\partial v_1}{\partial v_0}, \ldots, \frac{\partial v_n}{\partial v_0}\right)^T, \tag{7.46}$$

while $Q \in R^n$ is the vector with the components

$$Q_i := \frac{F_i^2}{a_0} \sum_{k=1,k\neq i}^{n} \frac{1}{v_k + a_k} > 0, \quad i = 1, \ldots, n. \tag{7.47}$$

Solving (7.45) for v'_{v_0} and making use of (7.47), one comes to the relationship

$$v'_{v_0} = -(I - J)^{-1} Q < 0, \tag{7.48}$$

which means that each component of $v\,(v_0, \tau)$ is a strictly decreasing function of $v_0 \geq 0$ for each fixed value of $\tau \in (-\infty, 0]$.

Now we are in a position to demonstrate the existence and smoothness of the unique solution $v(\tau) = (v_0(\tau), v_1(\tau), \ldots, v_n(\tau))$ of the complete system (7.34)–(7.35) for every fixed value of $\tau \in (-\infty, 0]$. To do that, we plug in the above-mentioned (uniquely defined for each fixed $\tau \in (-\infty, 0]$ and $v_0 \geq 0$) solution of subsystem (7.35) into (7.34) and arrive to the functional equation:

$$v_0 = \frac{1}{\sum_{i=1}^{n} \frac{1}{v_i\,(v_0, \tau) + a_i} - \tau}. \tag{7.49}$$

With the aim to prove the unique solvability of the latter equation, we fix an arbitrary $\tau \in (-\infty, 0]$ and introduce the function

$$\Psi\,(v_0) := \frac{1}{\sum_{i=1}^{n} \frac{1}{v_i\,(v_0, \tau) + a_i} - \tau}. \tag{7.50}$$

Since we know that

$$0 \leq v_i\,(v_0, \tau) \leq \frac{s}{n-1}, \quad n = 1, \ldots, n, \quad \text{where} \quad s = \max\{a_0, a_1, \ldots, a_n\}, \tag{7.51}$$

it brings us to the chain of relationships

$$\Psi\left(v_0\right) \leq \frac{1}{\sum_{i=1}^{n} \dfrac{1}{v_i\left(v_0, \tau\right)+a_i}} \leq \frac{1}{\sum_{i=1}^{n} \dfrac{1}{v_i\left(v_0, \tau\right)+s}}$$
$$\leq \frac{1}{\sum_{i=1}^{n} \dfrac{1}{\dfrac{s}{n-1}+s}} = \frac{1}{\sum_{i=1}^{n} \dfrac{n-1}{ns}} = \frac{s}{n-1}, \qquad (7.52)$$

which allows one to conclude that (for any fixed $\tau \in (-\infty, 0])$ the continuous function $\Psi = \Psi(v_0)$ maps the closed interval $\left[0, \dfrac{s}{n-1}\right]$ into itself. Therefore, according to Brouwer Fixed Point Theorem, there exists a fixed point $v_0 = \Psi(v_0)$ within this interval.

To finish the proof of the theorem, it is sufficient to establish that the above-determined fixed point is unique for each fixed $\tau \in (-\infty, 0]$ and, in addition, is monotone increasing with respect to τ. First, (7.48) implies that (for every fixed $\tau \in (-\infty, 0]$), the functions $v_i\left(v_0, \tau\right)$, $i = 1, \ldots, n$, are strictly decreasing by $v_0 \geq 0$; hence, each ratio $\dfrac{1}{v_i\left(v_0, \tau\right)+a_i}$, $i = 1, \ldots, n$, strictly increases with respect to $v_0 \geq 0$. Now we deduce from (7.53) below that the function $\Psi = \Psi(v_0)$, in its turn, strictly decreases with respect to $v_0 \geq 0$, which means that the fixed point $v_0 = \Psi(v_0)$ exists uniquely.

Differentiability of the thus determined well-defined function $v_0 = v_0(\tau)$ with respect to follows from Implicit Function Theorem, because

$$\frac{\partial \Psi}{\partial v_0} = \Psi^2 \sum_{i=1}^{n} \frac{\dfrac{\partial v_i}{\partial v_0}}{\left(v_i + a_i\right)^2} < 0, \quad \text{for any } \ \tau \in (-\infty, 0]. \qquad (7.53)$$

It is easy to see that the vector-function

$$v(\tau) := \left[v_0(\tau), v_1\left(v_0(\tau), \tau\right), \ldots, v_n\left(v_0(\tau), \tau\right)\right]^T$$

obtained by substituting the newly constructed function $v_0 = v_0(\tau)$ into the previously described solution of subsystem (7.35) represents the uniquely determined and continuously differentiable solution of the complete system (7.34)–(7.35).

In order to demonstrate the monotony of the above-described soluction's first component $v_0 = v_0(\tau)$ by τ, we differentiate equation (7.49) by the chain rule and make use of (7.50) to yield

$$\frac{dv_0}{d\tau} = \left[\Psi^2 \sum_{i=1}^{n} \frac{\dfrac{\partial v_i}{\partial v_0}}{\left(v_i + a_i\right)^2}\right] \cdot \frac{dv_0}{d\tau} + \Psi^2 \sum_{i=1}^{n} \frac{\dfrac{\partial v_i}{\partial \tau}}{\left(v_i + a_i\right)^2} + \Psi^2. \qquad (7.54)$$

Now solving (7.54) for the derivative $\frac{dv_0}{d\tau}$, one obtains:

$$\frac{dv_0}{d\tau} = \frac{B}{A}, \tag{7.55}$$

where, owing to (7.53),

$$A = 1 - \Psi^2 \sum_{i=1}^{n} \frac{\frac{\partial v_i}{\partial v_0}}{(v_i + a_i)^2} > 0 \tag{7.56}$$

while

$$B = \Psi^2 \left[\sum_{i=1}^{n} \frac{\frac{\partial v_i}{\partial \tau}}{(v_i + a_i)^2} + 1 \right] > 0, \tag{7.57}$$

according to (7.41). Combining (7.55)–(7.57), we conclude that $\frac{dv_0}{d\tau} = \frac{B}{A} > 0$; hence, the function $v_0 = v_0(\tau)$ strictly increases by τ. Moreover, by the evident estimate

$$v_0(\tau) \le -\frac{1}{\tau},$$

which follows from (7.49), we deduce that $v_0 = v_0(\tau)$ vanishes as $\tau \to -\infty$. Similarly, (7.43) implies that the same is valid for all the remaining influence coefficients, i.e., $v_i(\tau) \to 0$, $i = 1, \ldots, n$, as $\tau \to -\infty$. The proof of the theorem is complete. □

7.1.5 Numerical Results

In order to illustrate the difference between the mixed and classical oligopoly cases related to the conjectural variations equilibrium with consistent conjectures (influence coefficients), we apply formulas (7.21)–(7.22) to the simple example of an oligopoly in the electricity market from Liu et al. [206]. The only difference in our modified example from the instance of Liu et al. [206] is in the following: in their case, all six agents (suppliers) are private companies producing electricity and maximizing their net profits, whereas in our case, we assume that Supplier 0 (Supplier 5 in some instances) is a public enterprise seeking to maximize domestic social surplus defined by (7.4). All the other parameters involved in the inverse demand function $p = p(G)$ and the producers' cost functions, are exactly the same.

So, following Liu et al. [206], we select the IEEE 6-generator 30-bus system to illustrate the above analysis. The inverse demand function in the electricity market is given in the form:

Table 7.1 Cost functions' parameters

Agent i	b_i	a_i
0	2.00000	0.02000
1	1.75000	0.01750
2	3.00000	0.02500
3	3.00000	0.02500
4	1.00000	0.06250
5	3.25000	0.00834

$$p(G, D) = 50 - 0.02(G + D) = 50 - 0.02 \sum_{i=0}^{n} q_i. \tag{7.58}$$

The cost functions parameters of suppliers (generators) are listed in Table 7.1. Here, agents $0, 1, \ldots, 5$ will be combined in different ways in the examples listed below. In particular, Oligopoly 1 is the classical oligopoly where each agent 0–5 maximizes its net profit; Oligopoly 2 will involve agent 0 (public one, who maximizes domestic social surplus) and $1, \ldots, 5$ (private), whereas Oligopoly 3 comprises agents 5 (public) and $0, 1, \ldots, 4$ (private).

To find the consistent influence coefficients in their classical oligopoly market (Oligopoly 1), Liu et al. [206] use formulas (7.23) for all six suppliers, whereas for our mixed oligopoly model (Oligopoly 2), we exploit formula (7.21) for Supplier 0 and formulas (7.22) for Suppliers 1 through 5. With the thus obtained influence coefficients, the (unique) equilibrium is found for each of Oligopolies 1 and 2. The equilibrium results (influence coefficients, production outputs in MWh, equilibrium price, and the objective functions' optimal values in \$ per hour) are presented in Table 7.2 through 7.6. To make our conjectures v_i comparable to those used in Kalashnykova et al. [182], Kalashnikov et al. [165], and Liu et al. [206], we divide them by $[-p'(G)] = 0.02$ and thus come to $w_i := -v_i/p'(G) = 50v_i$, $i = 0, 1, \ldots, n$, shown in all the tables.

As it could be expected, the equilibrium price in Oligopoly 1 (classical oligopoly) turns out to be equal to $p_1 = 10.4304$, whereas in Oligopoly 2 (mixed oligopoly), it

Table 7.2 Computation results in consistent CVE: w_i, generation, profits

Agent i	w_i		q_i (MWh)		Profits (\$/h)	
	Oligopoly 1	Oligopoly 2	Oligopoly 1	Oligopoly 2	Oligopoly 1	Oligopoly 2
0	0.19275	0.18779	353.405	626.006	1727.4	595.77
1	0.19635	0.16674	405.120	358.138	2076.6	1550.04
2	0.18759	0.15887	258.463	220.451	1082.9	761.90
3	0.18759	0.15887	258.463	220.451	1082.9	761.90
4	0.17472	0.14761	142.898	125.462	707.5	538.37
5	0.22391	0.19270	560.180	488.905	2709.8	1917.98

drops down to $p_2 = 9.2118$. On the contrary, the total electricity power generation is higher: $G_2 = 2039.412$ MWh—in the second case (mixed oligopoly), than in Oligopoly 1, which is $G_1 = 1978.475$ MWh. Both results are more attractive for consumers. Simultaneously, the private producers' net profit values are observed to be lower in the mixed oligopoly (Oligopoly 2) than those in the classical oligopoly (Oligopoly 1.) In Oligopoly 2, profit is minimal in the cell of Agent 0, because its real objective function is *not* the net profit but *domestic social surplus* defined by (7.4); in this instance, it happens to reach $S = \$42,187.80/\text{h}$.

It is also interesting to compare the results in CVE with consistent conjectures (Oligopolies 1 and 2) against the production volumes and profits obtained for the same cases in the classical Cournot equilibrium (i.e., with all $w_i = 1, i = 0, 1, \ldots, 5$.) Table 7.3 illustrates the yielded results, with $p_1 = 14.760$ in the classical oligopoly (Oligopoly 1) much higher than the market equilibrium price $p_2 = 9.5349$ in the mixed oligopoly (Oligopoly 2).

Again, the total electricity power generation is higher: $G_2 = 2023.256$ MWh,—in the second case (mixed oligopoly), than in Oligopoly 1: $G_1 = 1761.9$ MWh. Both results are more propitious for consumers. Simultaneously, the private producers' net profit values are observed to be much lower in the mixed oligopoly (Oligopoly 2) than those in the classical oligopoly (Oligopoly 1). In Oligopoly 2, profit is even negative in the cell of Agent 0, as its objective function is not the profit but domestic social surplus defined by (7.4); in this example, it is equal to $S = \$35,577.50/\text{h}$. The latter data, together with the market price values, suggest that the mixed oligopoly with consistent conjectures is preferable to consumers than the Cournot model.

Table 7.3 Computation results in the Cournot models: w_i, generation, profits

Agent i	w_i		q_i (MWh)		Profits ($/h)	
	Oligopoly 1	Oligopoly 2	Oligopoly 1	Oligopoly 2	Oligopoly 1	Oligopoly 2
0	1.00000	1.00000	319.060	1200.000	3054.0	−5358.14
1	1.00000	1.00000	347.000	207.597	3461.7	1239.02
2	1.00000	1.00000	261.390	145.220	2220.5	685.38
3	1.00000	1.00000	261.390	145.220	2220.5	685.38
4	1.00000	1.00000	166.820	103.453	1426.2	548.51
5	1.00000	1.00000	406.230	221.767	3988.5	1188.70

Table 7.4 Computation results in the perfect competition model: w_i, generation, profits

Agent i	w_i		q_i (MWh)		Profits ($/h)	
	Oligopoly 1	Oligopoly 2	Oligopoly 1	Oligopoly 2	Oligopoly 1	Oligopoly 2
0	0.00000	0.00000	348.43	348.43	1214.00	1214.00
1	0.00000	0.00000	412.49	412.49	1488.80	1488.80
2	0.00000	0.00000	238.74	238.74	712.47	712.47
3	0.00000	0.00000	238.74	238.74	712.47	712.47
4	0.00000	0.00000	127.50	127.50	507.98	507.98
5	0.00000	0.00000	685.68	685.68	1960.50	1960.50

Of course, the perfect competition model (see, Table 7.4) with $w_i = v_i = 0, i = 0, 1, \ldots, 5$, is the best for consumers in both Oligopoly 1 and 2: with $p_1 = p_2 = 8.9685$ and the total produce $G_1 = G_2 = 2051.57$ MWh. Domestic social surplus is also higher in this case than in all the previous ones: $S = \$43,303.52/\text{h}$.

It is curious to note (cf. Tables 7.2, 7.3 and 7.4) that in the classical oligopoly (Oligopoly 1), the Cournot model demonstrates to be the most profitable for the producers, whereas it is not the case for the mixed oligopoly: here, the existence of a public enterprise with domestic social surplus as its utility function makes the consistent CVE more beneficial for the rest of suppliers than the Cournot one (except for the weakest Agent 4, for which, on the contrary, the Cournot model is most gainful).

Finally, we may compare the consistent CVEs (Table 7.5), Cournot equilibria (Table 7.6) and the perfect competition for the above-defined Oligopoly 2 (mixed oligopoly with Agent 0 being a public company) against a similar Oligopoly 3, in which not Agent 0 but the (much stronger) Agent 5 is the public supplier.

With the market price $p_3 = 7.8751$ even lower and domestic social surplus $= \$44,477.30/\text{h}$ even higher than those in the perfect competition model, this consistent CVE may serve as a good example of the strong public company realizing the implicit price regulation within an oligopoly.

A bit curious are the results reflected in Table 7.6: comparing the Cournot oligopoly in Oligopolies 1, 2, and 3, one may see that with a weaker public firm (Oligopoly 2), the private producers may incline to the Cournot model of behavior (cf. Table 7.3). However, with a stronger public supplier, as it is in Oligopoly 3,

Table 7.5 Computation results in consistent CVE: w_i, generation, profits

Agent i	w_i		q_i (MWh)		Profits ($/h)	
	Oligopoly 2	Oligopoly 3	Oligopoly 2	Oligopoly 3	Oligopoly 2	Oligopoly 3
0	0.18779	0.13208	626.006	259.480	595.77	851.16
1	0.16674	0.13497	358.138	303.229	1550.04	1052.75
2	0.15887	0.12803	220.451	176.884	761.90	471.22
3	0.15887	0.12803	220.451	176.884	761.90	471.22
4	0.14761	0.11843	125.462	105.984	538.37	377.63
5	0.19270	0.21584	488.905	1083.785	1917.98	114.52

Table 7.6 Computation results in the Cournot models: w_i, generation, profits

Agent i	w_i		q_i (MWh)		Profits ($/h)	
	Oligopoly 2	Oligopoly 3	Oligopoly 2	Oligopoly 3	Oligopoly 2	Oligopoly 3
0	1.00000	1.00000	1200.000	122.612	−5358.14	451.01
1	1.00000	1.00000	207.597	137.452	1239.02	543.18
2	1.00000	1.00000	145.220	86.766	685.38	244.67
3	1.00000	1.00000	145.220	86.766	685.38	244.67
4	1.00000	1.00000	103.453	71.569	548.51	262.51
5	1.00000	1.00000	221.767	1649.612	1188.70	−5319.04

private companies would rather select the consistent CVE: in the Cournot model, the strong public company subdues all the rivals to the minimum levels of production and profits. Nevertheless, the Cournot model with stronger public firm provides for the very low market price: $p_3 = 6.9045$, even though at the cost of a somewhat lower domestic social surplus: $S = \$41,111.59/\text{h}$.

As it could be expected, in the perfect competition model, both Oligopolies 2 and 3 give exactly the same results, albeit different domestic social surplus values: $S = \$43,303.52/\text{h}$ in Oligopoly 2 against a bit higher $S = \$44,050.04/\text{h}$ in Oligopoly 3 with the stronger public company.

In this section, we have studied a model of mixed oligopoly with conjectural variations equilibrium (CVE). The agents' conjectures concern the price variations depending upon their production output's increase or decrease. We establish existence and uniqueness results for the conjectured variations equilibrium (called an exterior equilibrium) for any set of feasible conjectures. To introduce the notion of an interior equilibrium, we develop a consistency criterion for the conjectures (referred to as influence coefficients) and prove the existence theorem for the interior equilibrium (understood as a CVE with consistent conjectures). To prepare the base for the extension of our results to the case of non-differentiable demand functions, we also investigate the behavior of the consistent conjectures in dependence upon a parameter representing the demand function's derivative with respect to the market price.

An interesting methodological question also arises: can the mixed oligopoly be related to collaborative game theory? Formally speaking, the mixed oligopoly is rather a cooperative than competitive game, as the public company's and the private firms' interests are "neither completely opposed nor completely coincident" (Nash [248]). At first glance, collaboration can be a worthwhile strategy in a cooperative game. However, according to Zagal et al. [327], "because the underlying game model is still designed to identify a sole winner, cooperative games can encourage anti-collaborative practices in the participants. Behaving competitively in a collaborative scenario is exactly what should not happen in a collaborative game".

7.2 Toll Assignment Problems

One of the most important problems concerning the toll roads is the setting of an appropriate cost for traveling through private arcs of a transportation network. In the section this problem is considered by stating it as a bilevel optimization (BLP) model. At the upper level, one has a public regulator or a private company that manages the toll roads seeking to increase its profits. At the lower level, several companies-users try to satisfy the existing demand for transportation of goods and/or passengers, and simultaneously, to select the routes so as to minimize their travel costs. In other words, what is sought is a kind of a balance of costs that bring the highest profit to the regulating company (the upper level) and are still attractive enough to the users (the lower level).

With the aim of providing a solution to the bilevel optimization problem in question, a direct algorithm based on sensitivity analysis is proposed in this section.

In order to make it easier to move (if necessary) from a local maximum of the upper level objective function to another, the well-known “filled function” method is used. Most results in this section are taken from Kalashnikov et al. [167].

7.2.1 Introduction

During the early years of industrial development, the production facilities were established near the consumers because the transportation was expensive, time-consuming, and risky. When transportation systems appeared, they allowed the producer to compete in distant markets, promoting economies of scale by increasing sales volume.

Due to the complexity of products and globalization, supply and distribution chains have grown enormously, therefore, logistics costs have “rocketed up” sharply. According to the data from the IMF (International Monetary Fund), logistics costs represent 12 % of gross national product, and they range from 4 to 30 % of the sales at the enterprise level.

Because of this growth, many countries have attached great importance to the development and modernization of the infrastructure to achieve greater participation in the global economy. There are organizations that deal with the development of communications and transportation infrastructure, building technologies to increase the coverage, quality and competitiveness of the infrastructure. In Mexico, administration of new (private) roads is commonly conceded to private companies, state governments, or financial institutions (banks, holdings, etc.), who set transportation tolls in order to retrieve money from the road users.

It has been recently noticed that under the concession model, there is less traffic flow using these tolled highways. One of the strategies taken to increase the use of toll roads is the regulation of tolls (pass rates). However, what are the appropriate criteria to assign these toll rates?

The problem here is how to assign optimal tolls to the arcs of a multicommodity transportation network. The toll optimization problem (TOP) can be formulated as a bilevel mathematical program where the upper level is managed by a firm (or a public regulator) that raises revenues from tolls set on (some) arcs of the network, and the lower level is represented by a group of users traveling along the cheapest paths with respect to a generalized travel cost. The problem can be interpreted as finding equilibrium among tolls generating high revenues and at the same time being attractive for the users. Other possible aims of the upper level decision maker can be found in Heilporn et al. [141], Didi-Biha et al. [88], Labbé et al. [196].

The problem in question has been extensively studied. In what follows, a literature review related to the TOP is made. Almost thirty years ago, Magnanti and Wong [212] presented a very complete theoretical basis for the uses and limitations of network design based on integer optimization with several models and algorithms. This provided a unification of network design models, as well as a general framework for deriving network design algorithms. They noticed that researchers had been motivated to develop a variety of solution techniques such as linear approximation methods and the search of vertices adjacent to the lowest cost flow problem threated as a *network design problem* (NDP).

The network design issues were mentioned several years later by Yang and Bell [322], who also provided a brilliant survey of the existing literature in this area. They introduced the elasticity concept in travel demand and the reserve capacity notion of the network in the NDP, which allowed them to obtain a network design problem easier to solve when trying to maximize an appropriate objective function. Moreover, they proposed an approach to NDP involving mixed elections, i.e., simultaneously adding links and improving the capabilities. The latter approach allowed the use of formulas based upon the maximization of consumer surplus as the objective function of the NDP. The authors mentioned that the challenge remained to develop a global search algorithm that could guarantee the optimality of a solution derived with the computationally efficient manner mentioned in [212].

Bell and Iida [17] sought a unification of the theoretical analysis of transportation networks, focused primarily on the assignment of stochastic user equilibrium (SUE), estimating trip tables and network's reliability. They saw the network design as an extension of the analysis of the transportation network, where the control of traffic signals is made in terms of an NDP. The latter is considered a difficult task because of its nonconvex nature and the complexity of the networks. They mentioned that the NDP can be posed as a bilevel optimization problem, where the upper level focuses on the network design to maximize certain goals, whereas the lower level determines how users react to changes in the network. In their monograph, they presented two different approaches to solve the problem of network design, one is the iterative method of design-assignment, which is relatively simple in its application and appears to converge quickly. The other is an iterative algorithm based on sensitivity analysis, which usually consumes more computational time to converge. The authors conclude that both methods provide different local optima, with slight differences in the objective function but significantly distinct in the design structure. Finally, they mentioned that in order to have a more satisfactory approach, it is necessary to combine bilevel optimization tools (to find a local solution) and a probabilistic search method (for comparing local solutions using simulated annealing), to come to a global solution.

Marcotte [218] mentioned that the NDP mainly deals with the optimal balance either of the transportation, investment, or maintenance costs of the networks subject to congestion. The network's users behave according to Wardrop's first principle of traffic equilibrium. He also suggested that the NDP can be modeled as a multilevel mathematical optimization problem.

Mahler et al. [213] dealt with the problem of congestion in road networks represented by two problems, namely, estimation of the trip matrix and optimization of traffic signals. Both problems were formulated as bilevel programs with allocation of stochastic user equilibrium (SUE) as the problem of the lower level. The authors presented an algorithm that gives a solution to the bilevel problem of estimation of the trip matrix and optimization of signals, making use of the "logit-based" model of assignment SUE at the lower level. The algorithm used applies standard routines to estimate the matrix at every iteration, and SUE assignment to find the search direction. The authors stated that it had not been possible to demonstrate the convergence results in general; however, in case that the optimal solution can be found by direct search, they demonstrated that the algorithm is able to give a good approximation of the optimal solution.

Lawphongpanich and Hearn [199] also examined the problem of traffic congestion as a problem of fixing the toll through a formulation with static demands. They mentioned that this problem can be classified into two types: (i) the problem of the first best solution, in which all the arcs of the network are tolled, and (ii) the problem of the second best solution, where it is assumed that some roads may have tolls and others not, which does not permit them to get the maximum benefit. The authors noticed that the latter problem can be posed as a bilevel optimization model, or as a mathematical program with equilibrium constraints (MPEC). They used the results achieved for the MPEC to develop a formulation equivalent to the nonlinear optimization problem for the second best solution. The latter is done in order to establish the properties of the second-best solution, which are of a particular interest to transport economists, and in its turn, help develop another algorithm to solve the problem in the nonlinear optimization formulation.

The pricing of road systems has a long history in the literature of transportation economics, as mentioned by Morrison [246], who worked with a theoretical framework developed through the empirical evidence of viability in pricing and policy. One can also find this concept in the engineering and road planning literature, as described in Cropper and Oates [47], who talked about the implementation of environmental economics in environmental policy design road systems; they focused on reducing traffic congestion on the roads through pricing to reduce negative aspects such as pollution. Other authors who treated the problem of traffic congestion were Arnott et al. [4]; they mentioned that the allocation of a uniform toll significantly reduces this problem by taking into account parameters of time (i.e., alternating departure times for users).

Hearn and Ramana [140] worked over the definition and optimization of different objectives under a given set of tolls that promote optimal traffic systems. Shifting a focus, one finds Viton's paper [309], which makes a comparison between the viability of private toll roads and highways free to users. The concept of maximizing profits through an optimal toll system is examined by Beckmann in [13] and by Verhoef in [303].

As mentioned before, bilevel optimization offers a convenient framework for modeling the optimal toll problems, as it allows one to take into account the user's behavior explicitly. Unlike the aforementioned investigations, Labbé et al. [197] considered the TOP as a sequential game between the owners of road systems (the leaders) and road users (the followers), which fits the scheme of a bilevel optimization problem. This approach is also implemented by Brotcorne [30] on the problem of fixing tariffs on cargo transportation. In the latter case, the leader is formed by a group of competing companies, and their earnings are determined by the total profits from the rates, while the follower is a carrier who seeks to reduce travel costs, taking into account the tolls set by the leader.

One of the simplest instances was analyzed by Kalashnikov et al. [173], where a TOP defined over a congestion-free, multicommodity transportation network was considered. In this setting, a highway authority (the leader) sets tolls on a subset of arcs of the network, while the users (the follower) assign themselves to the shortest (in terms of a generalized time) paths linking their respective origin and destination nodes. The goal of the leader in this context is to maximize the toll revenue. Hence, it is not in its interest to set very high tolls, because in this case the users would be discouraged from using the tolled sub-network. The problem can be stated as a combinatorial program that subsumes NP-hard problems, such as the Traveling Salesman Problem (see, Labbé et al. [196], for a reduction method). Following the initial NP-hardness proof in [196], computational complexity and approximation results were obtained by Marcotte et al. [220].

On the other hand, Dempe et al. [58] studied this problem designing a "fuzzy" algorithm for the TOP. Next, Lohse and Dempe [69] based their studies on the analysis of an optimization problem in some sense reverse to the TOP. In addition, Didi-Biha et al. [88] developed an algorithm based on the calculation of lower and upper bounds to determine the maximum gain from the tolls on a subset of arcs of a network transporting various commodities.

Studies have been conducted with roads without congestion and capacity limits, where it is assumed that congestion is affected by the introduction of tolls. This radically changes the mathematical nature of the model, and algorithms use a different approach. Such a model was presented by Yan and Lam in [321], but these authors were limited only to a simple model with two arcs. A more extensive work on the assumption of limited capacity arcs is presented by Kalashnikov et al. in [166], which studied four different algorithms to solve this problem.

The group of authors Brotcorne et al. [34] started investigating a bilevel model for toll optimization on a multicommodity transportation network as long ago as 2001. Recently, Brotcorne et al. analyzed this problem in [32] with the difference in that they allowed subsidies in the network; that is, they considered the tolls without constraints. The authors designed an algorithm that generated paths and then formed columns for determining the optimal toll values for the current path (the lower bound). Thereafter, they adjusted the revenue upper bound and finally applied a diversification phase. Also they validated their algorithm by conducting numerical experimentation and concluded that the proposed algorithm efficiently works in networks with few toll arcs. The same authors continued their work on the same problem in [31]. In the latter paper, they designed and implemented a tabu search algorithm, and concluded that their heuristics had obtained better results than other combinatorial methods. Dempe and Zemkoho [81] also studied the TOP and proposed a reformulation based on the optimal value function. This restatement has advantage over the KKT reformulation because it keeps on the information about the congestion in the network. They obtained optimality conditions for this restatement and established some theoretical properties for it.

The aim of the present section is to propose an algorithm based on the *allowable ranges to stay optimal* (ARSO) resulting from sensitivity analysis after solving the lower level problem. With this powerful tool, one can analyze possible changes in the coefficients of some variables in the objective function which do not affect the optimal solution (cf. the region of stability in Sect. 3.6.2.1). It also permits one to examine the effects on the optimal solution when the parameters take new values beyond the ARSO. This work is inspired by the previous research undertaken by Roch et al. [271].

In addition to dealing with the allowable ranges, the proposed technique also uses the concept of a “filled function” (see Renpu [267], Wan et al. [312], Wu et al. [318]), which is applied under the assumption that a local maximum (in our case) has been found. Then the “filled functions” technique helps one either to find another local maximum, better than the previous ones, or to determine that we have found (approximately) a best feasible or an optimal solution, according to certain parameters of tolerance.

The validity and reliability of this technique are illustrated by the results of numerical experiments with test examples used to compare the proposed approach with the other ones. Finally, the numerical results also confirmed the robustness of the presented algorithm.

To resume, in this section we propose and test two versions of a heuristic algorithm to solve the Toll Optimization Problem (TOP) based upon sensitivity analysis for linear optimization problems. The algorithm makes use of a sensitivity analysis procedure for the linear optimization problem at the lower level, as well as of the “filled functions” technicalities in order to reach a global optimum when “jammed” at some local optimum. The two versions of the method differ only in the way of selecting a new toll vector, namely, by changing only one toll value at a time, or by varying several toll values applying the well-known 100 % rule of sensitivity analysis.

The proposed heuristics aim at filling in a gap in a series of numerical approaches to the solution of TOP problem listed in the Introduction. To our knowledge, no systematic attempts to apply the sensitivity analysis tools to the toll assigned problem have been made. Moreover, the combination of these powerful tools with the “filled functions” techniques brings forward some new global optimization ideas.

Numerical experiments with a series of small and medium-dimension test problems show the proposed algorithm’s robustness and reasonable convergence characteristics. In particular, while ceding in efficiency to other algorithms when solving small problems, the proposed method wins in the case of medium (higher dimensional) test models.

The rest of the section is organized as follows. Section 7.2.2 contains the model statement and the definition of parameters involved. Section 7.2.3 describes the algorithm to solve the toll optimization problem, with Sect. 7.2.3.1 presenting the algorithm’s structure, Sect. 7.2.3.2 justifying the reduction of the lower level equilibrium problem to a standard linear program, and Sect. 7.2.3.3 recalling the “filled functions” technique. Section 7.2.4 lists the results of numerical experiments obtained for several test problems. Supplementary material (Sect. 7.2.5) describes the data of all the test problems tested in the numerical experiments.

7.2.2 TOP as a Bilevel Optimization Model

The methodology developed to solve this problem takes the model proposed by Labbé et al. [196] as a basis. They proved that the TOP can be analyzed as a leader-follower game that takes place on a multicommodity network $G = (K, N, A)$ defined by a set of origin-destination couples K, a node set N and an arc set A. The latter is partitioned into the subset A_1 of toll arcs and the complementary subset A_2 of toll-free arcs. We endow each arc $a \in A$ with a fixed travel delay c_a representing the minimal unit travel cost. Each toll arc $a \in A_1$ also involves a toll component t_a, to be determined. The latter is also expressed in time units, for the sake of consistency. The toll vector $t = \{t_a : a \in A_1\}$ is restricted by the vector $t^{max} = \left\{t_a^{max} : a \in A_1\right\}$ from above and by zero from below.

The demand side is represented by numbers n^k denoting the demand for transportation between the origin node $o(k)$ and the destination node $d(k)$ associated with commodity $k \in K$, $|K| = r$. A demand vector b^k is associated with each commodity. Its components are defined for every node i of the network as follows:

$$b_i^k = \begin{cases} n^k, & \text{if } i = d(k); \\ -n^k, & \text{if } i = o(k); \\ 0, & \text{otherwise.} \end{cases} \tag{7.59}$$

Let $x = \left\{x_a^k\right\}_{a \in A}$ denote the set of commodity flows along the arcs $a \in A$, and $\left\{i^+\right\} \subset A$ the set of arcs having i as their head (destination) node, while $\left\{i^-\right\} \subset A$ is the set of arcs having i as their tail (origin) node, for any $i \in N$. Based on the notation introduced above, the *toll optimization problem* (TOP) can be stated as the bilevel program (7.60)–(7.63):

$$\max_{t,x} F(t, x) = \sum_{k \in K} \sum_{a \in A_1} t_a x_a^k, \tag{7.60}$$

subject to

$$0 \le t_a \le t_a^{max}, \tag{7.61}$$

$$\forall k \in K \begin{cases} \varphi_k(t) = \min_{x^k} \left[\sum_{a \in A_1} (c_a + t_a)\, x_a^k + \sum_{a \in A_2} c_a x_a^k\right], \\ \textit{subject to} \\ \sum_{a \in \{i^+\}} x_a^k - \sum_{a \in \{i^-\}} x_a^k = b_a^k, \ \forall i \in N, k \in K, \\ x_a^k \ge 0, \ \forall a \in A, \ k \in K, \end{cases} \tag{7.62}$$

$$\sum_{k \in K} x_a^k \le q_a, \quad \forall a \in A. \tag{7.63}$$

In this (optimistic) formulation, both the toll and flow variables are controlled by the leader (the toll variables directly, the flow variables implicitly). On the other hand, the lower level constraints reflect the followers' intention to minimize their total "transportation costs", in terms of "time delay units" multiplied by the corresponding flow values, under current toll levels, and subject to the supply-demand requirements.

In order to prevent the occurrence of trivial situations, the following conditions are assumed in the same manner as in [88]:

1. For a certain amount of goods, demand from one node to another can be sent by arcs or paths that may be toll-free, depend on tolls, or combinations of both.
2. There is a transportation cost associated with each arc that is expressed as a cost at the lower level.
3. There is no profitable vector that induces a negative cost cycle in the network. This condition is satisfied if, for example, all delays c_a are nonnegative.
4. For each commodity, there exists at least one path composed solely of toll-free arcs.

7.2.3 The Algorithm

To find a solution of the TOP, we develop an algorithm dealing with the bilevel mathematical optimization model (7.60)–(7.63) starting from initial values t_a of tolls. With any toll vector fixed, we may treat the lower level problem as a linear program. After solving the latter by the simplex method, we perform sensitivity analysis for the lower level objective function. In the TOP analyzed here, the sum of the objective functions of all followers can be selected as the objective function in the lower level problem, see, Kalashnikov et al. [166]. If the analysis tells us that the current solution is a local maximum point for the upper level problem (this is so if sensitivity analysis does not allow to increase the coefficients of the basic flows along the toll arcs), we use the "filled functions" technique (described in Sect. 7.2.3.2; cf. e.g. Wu et al. [318, 319]) for the objective function of the leader. This allows us to make a "jump" to a neighborhood of another possible local maximum point, if the latter exists.

Once we have a new set of tolls, we proceed to solve the problem of the followers again and perform sensitivity analysis. If that does not allow more increases, we use the "filled functions" method again.

This procedure allows one to get an increase in the toll if the next local maximum is better; otherwise, after several fruitless attempts in a row, we stop with the last solution as approximately optimal.

7.2.3.1 Description of the Heuristic Algorithm

In this algorithm, we are going to combine the main structure of the method described by Kalashnikov et al. [173] and a new idea consisting in the following: A direct procedure may be represented as determination of the "fastest increase" direction for the upper level objective function in terms of the toll variables variations. The "formal gradient" of this objective function F from (7.60) can be determined by the current total flows along the toll arcs:

$$\frac{\partial F}{\partial t_a}(t, x) = \sum_{k \in K} x_a^k, \quad \forall a \in A_1. \tag{7.64}$$

We call it the "formal gradient" because the followers' optimal response is not taken into account in (7.64). However, as the fastest infinitesimal improvement direction, this vector can be used in our heuristic method. The possibility of solving a linear optimization problem at the lower level instead of the Nash equilibrium problem (7.60)–(7.63) has been justified in the papers [166, 173] by Kalashnikov et al.

In what follows, we present a description of the heuristic method proposed first by Kalashnikov et al. in [173] for solving the *congestion-free* case for the bilevel TOP, i.e., $q_a = +\infty$, $\forall a \in A$. However, the same algorithm can be also applied to solve the bilevel TOP problem with restricted capacities. This is justified by the following theoretical result.

7.2.3.2 A Simple Method to Solve a Special Generalized Nash Equilibrium Problem with Separable Payoffs

Consider a mapping $\Phi : X \to R^N$, where $X = X_1 \times X_2 \times \cdots \times X_N$ is a direct product of m subsets of Euclidean spaces: namely, $X_i \subset R^{n_i}$, $i = 1, \ldots, N$. Assume that the mapping Φ is *separable* in the sense that each of its components is restricted to its own domain, i.e., $\Phi_i : X_i \to R$, $i = 1, \ldots, N$. In other words, no two components of the mapping Φ share common variables. Many applied problems boast the latter property: cf. for example, the lower level of the Toll Optimization Problem, namely, the (generalized) Nash equilibrium problem (7.62)–(7.63).

Let us also consider two other mappings $G : X \to R^n$ and $H : X \to R^m$, which are *not necessarily* separable like the mapping Φ. Finally, let Ω be a subset of X defined as follows:

$$\Omega = \{x \in X : \; G(x) \le 0, \; H(x) = 0\}. \tag{7.65}$$

Now assume that we search for a *generalized Nash equilibrium* (GNE): Find a vector $x^* = \left(x_1^*, \ldots, x_i^*, \ldots, x_N^*\right) \in \Omega$ such that for every player $i = 1, \ldots, N$, the corresponding sub-vector $x_i^* \in X_i$ provides a point of a (global) *maximum* of its utility function (payoff) Φ_i over the subset $\Omega_i\left(x_{-i}^*\right) \subset X_i$ defined as follows:

$$\Omega_i\left(x_{-i}^*\right) = \left\{x_i \in X_i \text{ such that } \left(x_1^*, \ldots, x_{i-1}^*, x_i, x_{i+1}^*, \ldots, x_N^*\right) \in \Omega\right\}. \tag{7.66}$$

Here, $x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N) \in X_1 \times \cdots \times X_{i-1} \times X_{i+1} \times \cdots \times X_N$ is the complement to the vector $x_i \in X_i$ in the direct product X. In mathematical terms, what we seek is the following:

$$\Phi_i\left(x_i^*\right) = \max\left\{\Phi_i(x_i) : \; x_i \in \Omega_i\left(x_{-i}^*\right)\right\}, \quad \text{for all } \; i = 1, \ldots, N. \tag{7.67}$$

In what follows, we always suppose that

$$\Omega_i\left(x_{-i}\right) \neq \emptyset, \quad i = 1, \ldots, N, \tag{7.68}$$

for any $x \in \Omega$, i.e., each feasible solution of our GNE problem (7.65)–(7.67).

Now consider the following (scalar) mathematical optimization (MP) problem:

$$\varphi(x) \equiv \sum_{i=1}^{N} \Phi_i\left(x_i\right) \longrightarrow \max_{x \in \Omega}. \tag{7.69}$$

We are now in a position to state and prove the main result of this subsubsection:

Lemma 7.1 *Any solution of MP problem* (7.69) *is a generalized Nash equilibrium* (*GNE*)*, i.e., a solution of problem* (7.65)–(7.67).

Proof Assume that a vector $x^* = \left(x_1^*, \ldots, x_i^*, \ldots, x_N^*\right) \in \Omega$ solves problem (7.69). On the contrary, suppose that it is *not* an equilibrium for model (7.65)–(7.67). The latter means that for at least one player $i \in \{1, \ldots, N\}$, there exists another sub-vector $\bar{x}_i \in \Omega_i\left(x_{-i}^*\right)$ such that

$$\Phi_i\left(\bar{x}_i\right) > \Phi_i\left(x_i^*\right). \tag{7.70}$$

Now the mapping Φ being separable immediately implies the relationships

$$\begin{aligned}\varphi\left(\bar{x}\right) &= \sum_{j \neq i} \Phi_j\left(x_j^*\right) + \Phi_i\left(\bar{x}_i\right) > \sum_{j \neq i} \Phi_j\left(x_j^*\right) + \Phi_i\left(x_i^*\right) \\ &= \sum_{i=1}^{N} \Phi_i\left(x_i^*\right) = \varphi\left(x^*\right),\end{aligned} \tag{7.71}$$

where $\bar{x} = \left(x_1^*, \ldots, x_{i-1}^*, \bar{x}_i, x_{i+1}^*, \ldots, x_N^*\right) \in \Omega$. However, (7.71) means that $\varphi\left(\bar{x}\right) > \varphi\left(x^*\right)$, which contradicts the assumption that the above vector $x^* = \left(x_1^*, \ldots, x_i^*, \ldots, x_N^*\right) \in \Omega$ solves problem (7.69) and thus completes the proof. □

Remark 7.3 The result of Lemma 7.1 was obtained by Kalashnikov et al. in [166] in a bit more particular setting.

Now we return to the heuristic algorithm's description. Lemma 7.1 proved above permits one to justify Step 1 of the algorithm in question.

Algorithm: **`Step 1.`** `Set` $i = 0$. `Select` $t_a^{(i)} = t_a^{min} = 0$ `and minimize the aggregate lower level objective function`

$$h_{sum}(x) = \sum_{k \in K}\left[\sum_{a \in A_1}\left(c_a + t_a^{(i)}\right) x_a^k + \sum_{a \in A_2} c_a x_a^k\right], \tag{7.72}$$

subject to the flow conservation constraints and nonnegativity restrictions listed in (7.62) as well as the capacity constraints (7.63) in order to obtain the lower level's optimal response $x\left(t^{(i)}\right)$. Compute the leader's objective function value

$$F\left(t^{(i)}, x\left(\left(t^{(i)}\right)\right)\right) = \sum_{k\in K}\sum_{a\in A_1} t_a^{(i)} x_a^k. \tag{7.73}$$

If $i \geq 1$ then compare the upper level objective function value (7.73) with the same for the previous value of i, and if $F\left(t^{(i)}, x\left(\left(t^{(i)}\right)\right)\right) > F\left(t^{(i-1)}, x\left(\left(t^{(i-1)}\right)\right)\right)$ go to Step 2. Otherwise, go to Step 4. If this return to Step 4 from Step 1 occurs several times in a row (7 to 10), go to Step 5.

Step 2. Considering the *allowable ranges to stay optimal* (ARSO) given by the sensitivity analysis table obtained upon having solved the problem presented in Step 1, select the maximum increase parameters $\Delta_a^{k,+}$ for the (toll-arc) variables $x_a^k\left(t_a^{(i}\right)$, $a \in A_1$. Denote

$$A_1^+ = \left\{a \in A_1 : \sum_{k\in K} x_a^k\left(t_a^{(i}\right) > 0\right\}, \tag{7.74}$$

that is, the toll arcs with a positive current flow. According to (7.64) , these positive values are (nonzero) components of the "formal gradient" vector of the leader's objective function. If $A_1^+ = \emptyset$, then go to Step 4; otherwise, go to Step 3.

Step 3. The toll increment procedure can be implemented in two different ways. The first (more precautious) one consists in increasing the current toll value by the maximum allowable increment $\Delta_a^{k,+}$, $a \in A_1^+$, but not exceeding the corresponding component of the "formal gradient". More precisely, we set

$$t_a^{(i+1)} = \begin{cases} \min\left\{t_a^{max}, t_a^{(i)} + \max_{k\in K}\min\left\{\sum_{m\in K} x_a^m\left(t_a^{(i}\right), \Delta_a^{k,+}\right\}\right\}, & \text{if } a \in A_1^+; \\ t_a^{(i)}, & \text{otherwise.} \end{cases} \tag{7.75}$$

The second mode of computing the toll increment is determined by the combination of the allowable increase values:

$$t_a^{(i+1)} = \begin{cases} \min\left\{ t_a^{max}, t_a^{(i)} + \sum_{k\in K} \beta_k \min\left\{ \sum_{m\in K} x_a^m\left(t_a^{(i)}\right), \Delta_a^{k,+} \right\} \right\}, & \text{if } a \in A_1^+; \\ t_a^{(i)}, & \text{otherwise.} \end{cases} \tag{7.76}$$

Here, the nonnegative coefficients $\beta_k \geq 0$, $k \in K$, and such that $\sum_{k\in K} \beta_k = 1$, can be selected by the well-known 100-percent rule of sensitivity analysis. Next, if $t_a^{(i+1)} > t_a^{(i)}$ for at least one $a \in A_1^+$, then update $i := i + 1$ and close the loop by returning to Step 1 to minimize the lower level aggregate objective function with the updated toll values. Otherwise, i.e., if no toll has been increased, go to Step 4.

Step 4. The current set of tolls $\left\{t_a^{(i)}\right\}_{a\in A_1}$ apparently provides for a local maximum of the leader's objective function. In order to try to "jump" to some other possible local maximum solution, apply the "filled functions" technique described briefly in the next subsection. Then return to Step 1 and minimize the lower level aggregate objective function with the updated toll values.

Step 5. If, after a number of Steps 4 repeated (in our numerical experiments, we accepted this number as 7 to 10), one cannot improve the leader's objective function value, stop the algorithm, report the current vectors $\left\{t_a^{(i)}\right\}_{a\in A_1}$ and $x\left(t^{(i)}\right)$ as an approximation of a global optimum solution.

7.2.3.3 Application of the "Filled Functions" Technique

Our heuristic algorithm based upon sensitivity analysis also involves application of the "filled function" technique first proposed by Renpu [267]. This method works, according to the studies in [267], under the assumption that a local minimum of a function, which is continuous and differentiable in R^n, has been found. So the aim is to find another (better than the current) local minimum or determine that this is the global minimum of the function within the closed (polyhedral) constraint set $T \subset R^n$. Renpu [267] and Wu et al. [318, 319] defined "filled functions" for a minimization problem. Here, we adapt several versions of the "filled function" definitions and properties to deal with a maximization problem. For simplicity we assume that any local maximum point of the objective function has a positive value. Of course, the procedure is easily extended to the case where the value of the objective function can be negative, too.

Definition 7.3 Let $\bar{t}_0 \in T$ satisfy $\bar{t}_0 \neq t^*$ and $f\left(\bar{t}_0\right) \geq \frac{3}{4} f\left(t^*\right)$. A continuously differentiable function $P_{t^*}(x)$ is said to be a "filled function" for the maximization problem $\max_{t\in T} f(t)$ at a point t^* with $f\left(t^*\right) > 0$, if

1. t^* is a strict local minimizer of $P_{t^*}(t)$ on T;
2. any local maximizer $\bar{t}$ of $P_{t^*}(t)$ on T satisfies $f\left(\bar{t}\right) > \frac{3}{2} f(t^*)$ or $\bar{t}$ is a vertex of T;
3. any local maximizer $\hat{t}$ of the optimization problem $\max_{t\in T} f(t)$ with $f\left(\hat{t}\right) \geq \frac{7}{4} f(t^*)$ is a local maximizer of $P_{t^*}(t)$ on T;
4. any $\tilde{t} \in T$ with $\nabla P_{t^*}\left(\tilde{t}\right) = 0$ implies $f\left(\tilde{t}\right) > \frac{3}{2} f(t^*)$.

Now, define two auxiliary functions as follows: For any $d = f\left(t_k^*\right) > 0$, and $w = f(t)$, let

$$g_d(w) = \begin{cases} 1, & \text{if } w \geq \frac{1}{2}d; \\ 5 - \frac{48}{d}w + \frac{144}{d^2}w^2 + \frac{128}{d^3}w^3, & \text{if } \frac{1}{4}d \leq w < \frac{1}{2}d; \\ 0, & \text{if } w < \frac{1}{4}d, \end{cases} \tag{7.77}$$

and

$$h_d = \begin{cases} w - \frac{1}{4}d, & \text{if } w \leq \frac{1}{4}d; \\ \left(\frac{16}{d^2} - \frac{128}{d^3}\right)w^3 + \left(\frac{144}{d^2} - \frac{20}{d}\right)w^2 + \left(8 - \frac{48}{d}\right)w + 5 - d, & \text{if } \frac{1}{4}d < w \leq \frac{1}{2}d; \\ 1, & \text{if } \frac{1}{2}d < w \leq \frac{3}{2}d; \\ -\frac{128}{d^3}w^3 + \frac{624}{d^2}w^2 - \frac{1008}{d}w + 541, & \text{if } \frac{3}{2}d < w \leq \frac{7}{2}d; \\ 2, & \text{if } w > \frac{7}{4}d. \end{cases} \tag{7.78}$$

Given a $t^* \in T$ such that $f\left(t^*\right) > 0$, we define the following "filled function":

$$G_{q,t^*}(t) = -\exp\left(-\|t - t^*\|^2\right) g_{\frac{f(t^*)}{4}}\left(f(t)\right) - q h_{\frac{f(t^*)}{4}}\left(f(t)\right), \tag{7.79}$$

where $q > 0$ is a parameter. This "filled function" will be used in our algorithm.

First, based on Wu et al. [318] we have the following result:

Theorem 7.4 *Assume that the function $f : R^n \to R$ is continuously differentiable and there exists a polyhedron $T \subset R^n$ with $t_0 \in T$ such that $f(t) \leq \frac{1}{2} f\left(t_0\right)$ for all $t \in R^n \backslash int\ T$. Let $\bar{t}_0 \neq t^*$ be a point such that $f\left(t^*\right) - \left(\bar{t}_0\right) \leq \frac{1}{4} f\left(t^*\right)$. Then:*

1. *there exists a $q^1_{t^*} \geq 0$ such that when $q > q^1_{t^*}$, then any local maximizer $\bar{t}$ of the mathematical program $\max_{t\in T} G_{q,t_0^*}(t)$ obtained via the search starting from $\bar{t}_0$ satisfies $\bar{t} \in int\ T$;*
2. *there exists a $q^2_{t^*} > 0$ such that if $0 < q \leq q^2_{t^*}$, then any stationary point $\tilde{t} \in T$ with $\tilde{t} \neq t^*$ of the function $G_{q,t_0^*}(t)$ satisfies $f\left(\tilde{t}\right) > \frac{3}{2} f\left(t^*\right)$.*

Proof The proof is almost identical to that of Theorem 2.2 in Wu et al. [318]. □

Making use of the auxiliary function (7.79) we can detail the "jump" to a neighborhood of another local maximum point of the upper level objective function F.

Algorithm: Step 1. Let our current toll iteration $t^{(i)}$ be such that formulas (7.75) and (7.76) provide no increase in the toll values. It can be shown that maximization of the auxiliary function (7.79) instead of the original upper level function F is equivalent to a (moderate) increase of the toll parameters $t^{(i)}$ (one or several of them, depending on the mode applied: (7.75) or (7.76)).

Step 2. If the new optimal response $x\left(t^{(i+1)}\right)$ is related to new ARSO upper bounds distinct from zero, return to Step 1 of the algorithm and continue increasing the toll parameters according to formulas (7.75) or (7.76).

Step 3. Otherwise, i.e., if the new ARSO upper bounds are all zero, double the increment of the toll parameters and return to Step 2. If this happens several times without success (i.e., the ARSO upper bounds continue to be zero), go to Step 5 and finish the computational algorithm.

After having defined the above procedures, we are going to illustrate the steps of the combined proposed sensitivity analysis-"filled function" algorithm to solve the TOP.

In Fig. 7.1, we begin by assigning initial values of zero toll cost. After solving the linear optimization problem of the follower to determine the flow in the arcs and obtaining a value for the leader's objective function, sensitivity analysis of the follower is performed, taking into account only toll-arc variables of the current solution. Then having listed the possible increases in the coefficients of the objective function of the follower derived from the sensitivity analysis data, and based upon the formal gradient vector of the upper level objective function F, we update the present toll vector $\left\{t_a^{(i)}\right\}_{a\in A_1}$. When positive increments of t cannot be obtained anymore based on sensitivity analysis and the formal gradient of the function F, apply the "filled function" procedure. A new function is created based on the leader's objective function and a new toll vector is probed. Once there is a new toll vector, go to Step 1 and close the loop. The algorithms stop if neither sensitivity analysis nor the "filled function" method provide a better value for the leader's objective function after several (say,

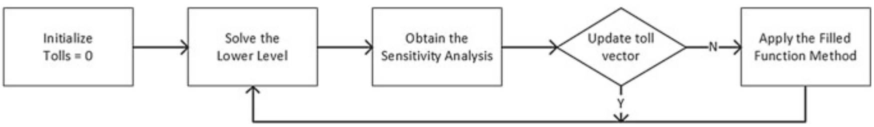

Fig. 7.1 Diagram of the combined method

7–10) attempts in a row, which can mean that an approximate global optimum has been reached, and the algorithm stops. The multicommodity flows corresponding to the final toll values give approximate optimal solutions for the follower, too.

7.2.4 Results of Numerical Experiments

In order to verify the performance of the algorithm, we conducted experiments on two different graphs, each with five different instances. In order to make valid comparisons of the efficiency and computational time of the proposed algorithm we emulated the experimentation conducted by Kalashnikov et al. [166] with their four different proposed algorithms. The following paragraphs describe the environment under which the experimentation was carried out and then describe the methodology used for the application of the algorithm.

In order to check the proposed heuristic sensitivity analysis algorithm combined with the method of "filled functions" (FF), a personal computer was used. The characteristics of the computer equipment used for the development and implementation of the algorithm were: Intel (R) Atom (TM) CPU N455 with a speed 2.00 GHz and 1.67 GB of RAM memory. The coding algorithm was written in the Matlab mathematical software in its version MatLab R2010a. This software was used due to its linear optimization tools in the "Optimization Toolbox". One of the functions used was `linprog` because the lower level of the TOP can be replaced by a corresponding linear optimization problem of the minimum cost flow.

The main parameters of the problems are the ones that define the size of the network: the number of nodes $|N|$, arcs $|A|$, toll arcs $|A_1|$, and commodities $|K|$. Each toll-free arc and toll arc has been assigned a fixed time-delay value c_a generated pseudo-randomly. The problems involved are of small size with two commodities. The graphs and the parameters of the tested instances can be found in the Supplementary material in Sect. 7.2.5. As mentioned above, the sizes of the networks were:

Network 1: 7 nodes, 12 arcs, of which 7 are toll arcs.
Network 2: 25 nodes, 40 arcs, of which 20 are toll arcs.

The results for each example can be seen in the Tables 7.7 to 7.18 below. The first column (called SA + FF) in each table shows the data related to the proposed algorithm, in which the increase in the tolls after sensitivity analysis is conducted in the first mode (cf. 7.75). Analogously, the second column (with the heading SA + FF 100 %) involves the results generated with the developed algorithm, updating the current tolls by formula (7.76). The next four columns show the results obtained after emulating the algorithms proposed in Kalashnikov et al. [166], that is, the Nelder-Mead (NM), Penalization (P), Quasi-Newton (QN), and Gradient (G) methods. The best obtained result is marked in bold.

Tables 7.7 and 7.8 may be a base for the assertion that the approximate solution obtained by all six methods applied to the test problems 1 and 2 are practically the same, which could mean that they are indeed the desired global maximum solutions for the leader.

The possible ways of measuring the algorithms efficiency are: to compare, first, the number of iterations required for each algorithm to reach an approximate solution for a given tolerance value, and second, the average computational cost (the number of iterations necessary on average) to decrease the error by one decimal order. This metric is calculated by the following formula:

$$Cost_{iter} = \frac{\#_{iter}}{\log_{10} \varepsilon_0 - \log_{10} \varepsilon_f}, \tag{7.80}$$

where $\#_{iter}$ denotes the number of iterations needed to reach the optimal value, ε_0 is the initial error computed as the difference between the initial leader's objective function value and the final one reached by the algorithm, this is, $\varepsilon_0 = |F_0 - F^*|$. In a similar manner, ε_f is the approximate final error calculated as the (absolute value of the) difference of the leader's objective function values evaluated at the last two approximate solutions. Tables 7.9 and 7.10 present the total number of iterations required for each algorithm, and Tables 7.11 and 7.12 shows the average cost of the number of iterations required to reduce the error by one order.

Tables 7.9 and 7.10 illustrate that the number of iterations the tested algorithms needed to reach approximately optimal solutions in both test sample problems have

Table 7.7 Leader's objective function value for Network 1

N1	SA + FF	SA + FF 100 %	NM	P	QN	G
1	161.9975	**162.9989**	162.9987	162.8215	162.9972	162.9134
2	274.9905	274.9979	**274.9996**	274.8320	274.9975	274.9321
3	57.98889	**58.9998**	58.9996	58.8719	58.9979	58.9229
4	**155.9806**	156.9980	156.9971	156.8504	156.9988	156.9057
5	136.9888	136.9984	136.9989	**136.8408**	136.9972	136.9750

Table 7.8 Leader's objective function value for Network 2

N2	SA + FF	SA + FF 100 %	NM	P	QN	G
1	1761.488	**1763.984**	1763.876	1762.887	1763.963	1762.629
2	2758.542	**2758.926**	2758.804	2758.237	2758.924	2758.484
3	2364.98	**2367.89**	2365.45	2365.98	2367.82	2365.33
4	3785.41	**3790.99**	3789.24	3790.11	3790.99	3790.18
5	610.99	**611.99**	611.91	610.91	611.97	611.43

Table 7.9 Number of iterations required to solve Network 1

N1	SA + FF	SA + FF 100 %	NM	P	QN	G
1	478	384	277	489	**4**	170
2	510	399	269	484	**8**	397
3	263	195	275	487	**18**	529
4	406	337	164	518	**13**	336
5	276	205	205	469	**10**	108

Table 7.10 Number of iterations required to solve Network 2

N2	SA + FF	SA + FF 100 %	NM	P	QN	G
1	592	**479**	1,412	787	581	745
2	636	546	1,587	685	**496**	812
3	734	411	1,464	633	**374**	596
4	586	497	1,286	549	**324**	893
5	556	418	1,698	591	**309**	650

Table 7.11 Average cost in the number of iteration to reduce the error for Network 1

N1	SA + FF	SA + FF 100 %	NM	P	QN	G
1	197.4995	131.8796	36.9995	348.9920	**6.3082**	41.9907
2	109.1328	80.7525	**38.8531**	284.6465	50.2147	122.6097
3	63.3016	55.5471	43.9853	275.8310	**18.9017**	185.9731
4	160.8014	168.6568	**26.0900**	236.0231	30.4811	108.9364
5	58.8175	42.5245	33.7262	301.8081	**29.4384**	35.7321

Table 7.12 Average cost in the number of iteration to reduce the error for Network 2

N2	SA + FF	SA + FF 100 %	NM	P	QN	G
1	440.6532	356.5420	915.1583	896.1183	285.9969	**239.4804**
2	266.6127	**228.8844**	876.7644	692.3814	277.0817	313.9064
3	465.8475	260.8492	774.8305	818.0137	**186.8402**	291.4508
4	314.3834	266.6357	464.7329	619.5010	**159.6143**	259.2267
5	326.4375	**245.4152**	944.1694	693.3167	360.4346	296.6317

the same order, with a single exception of the Nelder-Mead method. The latter is known to need more iterations in general. The Nelder-Mead method is a derivative-free algorithm, i.e., it does not use even the first derivatives of the upper level objective function.

Table 7.13 Number of objective function values evaluated to solve instances for Network 1

N1	SA + FF	SA + FF 100 %	NM	P	QN	G
1	1,563	1,174	3,249	492	**127**	4,506
2	1,633	807	3,860	487	**231**	10,347
3	544	398	3,948	**490**	507	13,713
4	1,821	1,495	2,367	521	**398**	7,993
5	565	416	2,939	472	**284**	2,142

It seems (from Tables 7.11 and 7.12) that our sensitivity analysis-based algorithms are quite competitive against the other methods when the dimension of the test problem is larger. Such robustness of the procedure may help when dealing with real-life problems, which are usually of larger dimensions.

In Tables 7.13 to 7.16, we also measured the number of values of the upper level objective function calculated during the performance of the algorithms and the average computational cost (measured in the number of objective function evaluations necessary to reduce the error by one decimal order). The evaluation formula used in Tables 7.15 and 7.16 is:

$$Cost_{obj} = \frac{\#_{obj}}{\log_{10} \varepsilon_0 - \log_{10} \varepsilon_f}, \tag{7.81}$$

where $\#_{obj}$ is the number of the leader's objective function evaluations until the algorithm stops.

Again, the proposed sensitivity analysis-based methods performed at a quite high level of efficiency compared to the best (quasi-Newton) algorithm even when the total number of objective function calculations is taken into account, but only for larger problems (see Table 7.14).

According to Tables 7.15 and 7.16, with respect to the *average cost* in the number of values of the leader's objective function calculated to reduce the order of error by 1, our sensitivity analysis-based methods performed better both for small and medium-sized test problems, which is a promising feature.

The last measure we checked in order to compare the algorithms' performance is the computational time that they needed to reach a good approximate solution. It is

Table 7.14 Number of objective function values evaluated to solve instances for Network 2

N2	SA + FF	SA + FF 100 %	NM	P	QN	G
1	4,717	**3,582**	15,224	17,257	4,545	18,311
2	4,840	**3,131**	12,996	14,037	6,183	63,752
3	5,312	**3,378**	9,873	16,779	5,797	49,937
4	4,454	**3,592**	8,486	12,534	4,644	73,227
5	4,210	**3,504**	8,094	13,736	5,292	51,781

Table 7.15 Average cost in the objective functions evaluations for the instances of Network 1

N1	SA + FF	SA + FF 100 %	NM	P	QN	G
1	396.7761	265.1142	529.5666	351.1331	**200.2873**	1113.0009
2	349.4392	**163.3266**	557.5213	286.4108	449.9496	3195.5751
3	130.9357	**112.8034**	631.4701	277.5302	532.3981	4820.8882
4	323.2002	342.1193	376.5560	**237.3900**	933.1916	2591.4543
5	120.4053	**86.2937**	483.5187	303.7387	836.0515	708.6885

Table 7.16 Average cost in the objective functions evaluations for the instances of Network 1

N2	SA + FF	SA + FF 100 %	NM	P	QN	G
1	1138.1844	**864.3155**	6969.9441	2013.1342	1813.1588	15966.7176
2	889.0858	**575.1503**	5949.9076	1403.6637	1876.7644	19564.7359
3	982.0206	**624.4852**	2820.4123	1397.8413	1774.8305	16041.0766
4	793.3171	**639.7833**	5228.1712	2474.2977	1730.7076	21256.8816
5	726.8940	**604.9967**	9441.2869	1151.3150	1810.6673	17110.1938

Table 7.17 Required computational time to solve the instances for Network 1 (in seconds)

N1	SA + FF	SA + FF 100 %	NM	P	QN	G
1	16.8807	13.4929	28.9564	23.4072	**1.2825**	136.8533
2	15.7290	12.2325	31.5437	24.6059	**5.7874**	131.3832
3	17.0548	13.5037	35.8244	22.2121	**6.9523**	111.3348
4	14.4474	11.7556	39.9163	24.4446	**5.3806**	145.5045
5	12.9696	10.2944	35.2639	23.2295	**3.6844**	124.4190

Table 7.18 Required computational time to solve the instances for Network 2 (in seconds)

N2	SA + FF	SA + FF 100 %	NM	P	QN	G
1	311.9787	**263.3140**	549.9019	297.6871	486.6073	657.8525
2	532.5229	444.5805	498.6304	306.3302	**185.2478**	627.73277
3	462.7594	**391.2001**	652.4052	588.3764	562.5629	595.5689
4	279.0327	238.9128	430.6348	255.2499	**133.1982**	751.8234
5	573.2400	**488.5338**	581.2160	575.3600	578.9995	507.3447

important to mention that we emulated the benchmark algorithms, so the required time is going to be valid because we have run all the experiments on the same computer. Tables 7.17 and 7.18 present the time (in seconds) used for each instance and each network.

The corresponding two Tables 7.17 and 7.18 again demonstrated that our algorithms ceded the leading position only to the quasi-Newton method that was proven to be extremely fast when applied to the low-dimensional problems. However, in the higher-dimensional examples, the sensitivity-analysis-based procedure didn't lag behind, even overwhelming all the other methods tested here.

7.2.5 Supplementary Material

In this supplementary material, we present the two networks considered during the experimentations described. In Figs. 7.2 and 7.3, the dotted lines denote the toll arcs, while the regular lines correspond to the toll-free arcs.

Also, we specify the parameters used in the two examples we solved in order to compare the algorithms' performance. Here, we list the travel costs c_a, the demands n^k, the commodities' origin-destination pairs $p = \{(o(k), d(k))\}_{k \in K}$, where $o(k)$ represents the origin node, and $d(k)$ denotes the destination node; $k \in K$, with $|K| = 2$. It is important to mention that in these experiments, we do not restrict the arc capacities.

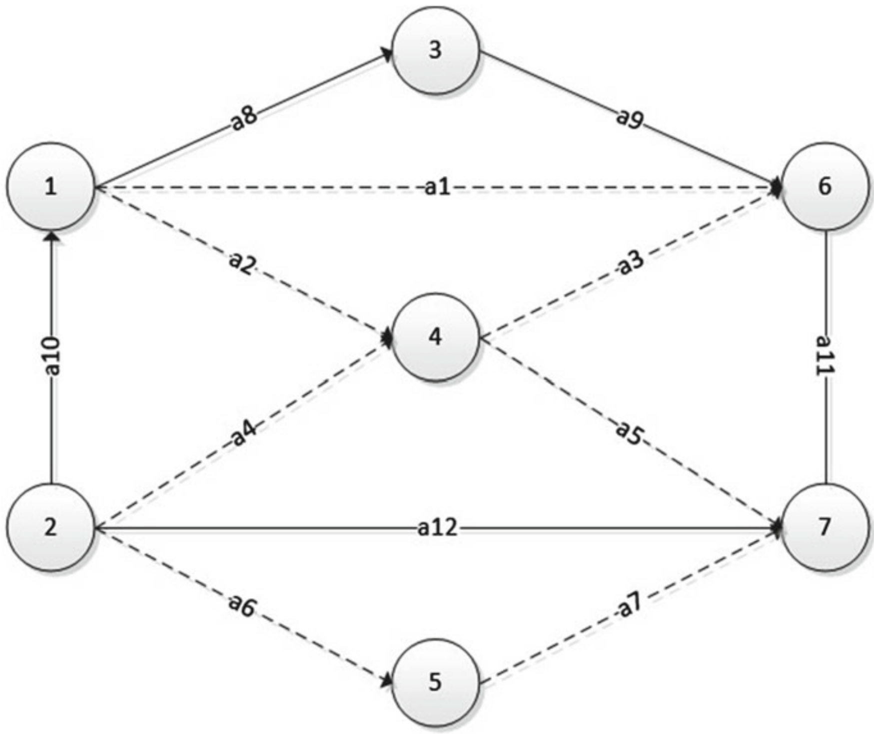

Fig. 7.2 Network 1

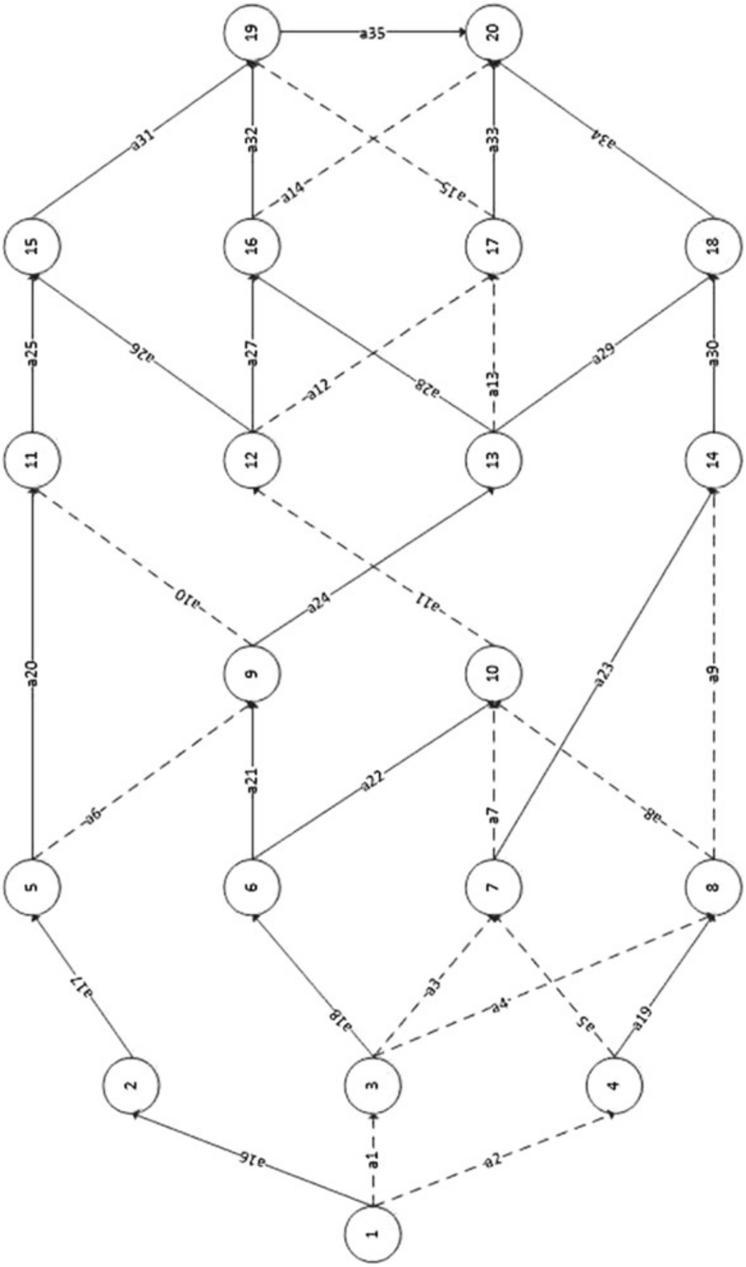

Fig. 7.3 Network 2

Table 7.19 Parameters considered in the instances for Network 1

NN of inst.	Parameters
1	$c = (1, 2, 5, 4, 3, 3, 2, 7, 4, 3, 8, 12);\ p = \{(1, 6), (2, 7)\};\ n^k = (10, 9)$
2	$c = (3, 4, 2, 2, 3, 3, 4, 9, 9, 5, 6, 15);\ p = \{(1, 6), (2, 7)\};\ n^k = (15, 5)$
3	$c = (4, 3, 2, 1, 1, 3, 2, 5, 6, 3, 1, 5);\ p = \{(1, 6), (2, 7)\};\ n^k = (5, 8)$
4	$c = (1, 3, 1, 2, 3, 1, 1, 5, 4, 2, 4, 13);\ p = \{(1, 6), (2, 7)\};\ n^k = (5, 12)$
5	$c = (3, 4, 5, 3, 3, 6, 2, 7, 7, 8, 10, 9);\ p = \{(1, 6), (2, 7)\};\ n^k = (10, 9)$

Table 7.20 Parameters considered in the instances for Network 2

NN of inst.	Parameters
1	$c = (1, 3, 4, 2, 1, 2, 2, 2, 2, 2, 4, 5, 1, 7, 9, 2, 4, 8, 7, 4, 4, 10, 12, 11, 11, 12, 9, 11, 4, 10, 9, 13, 16, 12, 10, 13, 12, 10, 7, 9);\ p = \{(1, 12), (2, 19), (2, 25)\};\ n^k = (12, 24, 30)$
2	$c = (9, 3, 7, 1, 5, 3, 4, 4, 4, 9, 1, 4, 6, 5, 6, 1, 6, 7, 7, 4, 6, 5, 2, 4, 7, 7, 8, 6, 10, 6, 5, 3, 8, 6, 11, 10, 9, 3, 5, 4);\ p = \{(1, 12), (2, 19), (1, 25)\};\ n^k = (31, 41, 120)$
3	$c = (4, 8, 1, 7, 3, 9, 5, 5, 2, 7, 6, 6, 4, 9, 5, 5, 9, 5, 1, 4, 9, 5, 1, 4, 9, 3, 9, 1, 8, 4, 6, 3, 9, 1, 1, 1, 2, 5, 1, 10);\ p = \{(2, 23), (2, 19), (1, 12)\};\ n^k = (48, 50, 31)$
4	$c = (1, 5, 2, 6, 3, 5, 2, 3, 7, 2, 5, 1, 6, 9, 3, 1, 3, 8, 1, 1, 10, 8, 9, 11, 6, 7, 10, 7, 2, 7, 7, 6, 9, 10, 6, 10, 5, 8, 5, 9);\ p = \{(1, 25), (2, 19), (2, 25)\};\ n^k = (84, 45, 71)$
5	$c = (4, 3, 6, 4, 4, 3, 2, 3, 3, 2, 7, 3, 4, 5, 7, 1, 6, 4, 4, 5, 7, 3, 5, 10, 10, 9, 10, 10, 10, 7, 7, 8, 11, 10, 10, 8, 8, 9);\ p = \{(1, 25), (2, 23), (2, 25)\};\ n^k = (10, 6, 8)$

First, we show the topology of Network 1 represented with a graph with 12 arcs and 7 nodes. For the two commodities transported within this network, we cite the parameters of the TOP problem.

The values of parameters used in the instances for Network 1 are listed in Table 7.19.

Finally, we describe Network 2, which consists in 25 nodes, 40 arcs and 3 commodities in Fig. 7.3. Here, again, the dotted lines are toll arcs, the regular lines represent toll-free highways. The values of parameters of the considered instances are collected in Table 7.20 (recall, that here $|K| = 3$).