

Eliciting Coordination with Rebates

Patrick Maillé

Institut Telecom, Telecom Bretagne, Université Européenne de Bretagne,
Cesson Sévigné, France, patrick.maille@telecom-bretagne.eu

Nicolás E. Stier-Moses

Graduate School of Business, Columbia University, Uris Hall 418,
New York, New York 10027, stier@gsb.columbia.edu

This article considers a mechanism based on rebates that aims at reducing congestion in urban networks. The framework helps select rebate levels so that enough commuters switch to modes that are under used. Indeed, getting a relative small number of drivers to switch to public transportation can significantly improve congestion. This mechanism is modeled by a Stackelberg game in which the transportation authority offers rebates, and participants factor them into the costs of each mode. A new Wardrop equilibrium arises when participants selfishly select one of the modes of transportation with the lowest updated costs. Rebate levels are chosen taking into account not only the potential reduction of the participants' cost, but also the cost of providing those rebates. Part of the budget for rebates may come from the savings that arise from the more efficient use of capacity. We characterize the Stackelberg equilibria of the game, and describe a polynomial-time algorithm to compute the optimal rebates for each mode. In addition, we provide tight results on the worst-case inefficiency of the resulting Wardrop equilibrium, measured by the so-called price of anarchy. Specifically, we describe the tradeoff between the sensitivity of the owner towards rebate costs and the worst-case inefficiency of the system.

Key words: network pricing; subsidies as incentives; Wardrop equilibrium; Stackelberg games; price of anarchy

History: Received: September 2008; revision received: March 2009; accepted: April 2009. Published online in *Articles in Advance* October 16, 2009.

1. Introduction

Congestion in most large cities in the world is prevalent. The Urban Mobility Study, a survey conducted by the Texas Transportation Institute (Schrank and Lomax 2007), estimated that the *congestion bill* related to automobile traffic, in the U.S. alone, amounts to \$78.2 billion in 2005. This cost estimate is based on the following two components: 4.2 billion hours of delay that people lose to highway congestion plus 2.9 billion gallons worth of fuel. Given those figures, even a small improvement in the efficiency of the road traffic system implies that a large sum of money and time could be saved. Furthermore, a recent study by the (Partnership for New York City 2006, p. 3) concluded that “traffic delays add to logistical, inventory, and personnel costs that annually amount to an estimated \$1.9 billion in additional costs of doing business and \$4.6 billion in unrealized business revenue.”

In most urban transportation networks, commuters do not have to pay the cost they impose to others by a particular choice of mode and route. Because of these under-payments, decisions—which are mostly influenced by a desire to get to the destination as fast as possible and as cheap as possible—lead to choices that do not use the available capacity of the network well. Because congestion increases sharply with road use,

having relatively few drivers switch to other modes significantly improves commute times. Starting with the seminal idea of Vickrey (1955, 1969), many transportation economists have advocated the use of congestion pricing to achieve this goal. The scheme forces drivers to pay a toll when entering congested areas as an incentive to switch to other modes of transportation (operational details differ according to the concrete implementation). The underlying idea is to charge drivers the externality they impose to others because when commuters internalize these externalities, the corresponding choices maximize the system welfare.

Singapore introduced congestion pricing in 1975, London in 2003 (Santos 2005; Santos and Fraser 2006), and Stockholm in 2007. Increasingly, many large cities have been debating whether a congestion pricing scheme should be adopted, New York City being the most prominent example in the United States. Nevertheless, it has been very hard to implement congestion pricing because of technical, economical, and political problems (e.g., the proposal in New York was not implemented after the State Assembly blocked it in 2008). Even though proponents claim it will decrease the delay costs generated by congestion, will curb harmful emissions and will reduce the dependence

on oil, the main concern that opponents raise comes from the perspective of social equity. Opponents favor other alternatives such as restricting some cars from driving to congested areas on some days of the week, increasing the taxes for parking, and offering incentives for tele-commuting, among many others.

Introducing congestion pricing is not likely to have a large impact among the better-positioned segment of society. They will continue to drive because they can afford to pay the corresponding charge. In contrast, the not-as-well positioned segment will be relegated to the less desirable options because they cannot afford to pay the tolls. Some articles suggest different measures to alleviate problems of inequity raised by this type of mechanisms (e.g., Starkie 1986; Button and Verhoef 1998).

The most important practical questions are which incentives to offer, at what level, how much they will improve social welfare, and who will be affected. Most cities that do use congestion pricing, use a *second-best* approach because charging tolls on each arc is not feasible or practical, even with electronic toll collection systems.¹ Besides the implementation cost, charging (potentially variable) tolls everywhere makes it more complicated for the driver to select a route. In the future, this may be less of a problem because the market penetration of route guidance devices is likely to be larger, and it is technologically feasible that these devices receive broadcasts with the current values of tolls. Most cities that have adopted congestion pricing decided to charge a flat daily fee that has to be paid on each day a driver wants to access the central business district of the city. Although a flat fee does not elicit the most efficient choices, it is conducive to increase the social welfare. Indeed, the high cost of the charge is enough to provide a detriment to some drivers who will switch to other modes of transportation. Unfortunately, an implementation of a congestion pricing scheme is not likely to allow for much room for experimentation. If not done right initially, expected benefits may not be realized, thereby invalidating the whole effort and potentially jeopardizing the political viability of a second try. Quantitative models can be used upfront to help policymakers make decisions and compare proposals.

This article initiates the study of an approach that complements congestion pricing. Although congestion pricing considers only (positive) tolls, there is no reason not to use negative tolls, which we refer to as

rebates or subsidies. Often, the proceeds of congestion pricing are used to fund improvements in public transportation, but very rarely are they used to reduce operating costs by subsidizing fares. On the contrary, it has been documented that in some cases public transportation fares increased after the introduction of congestion pricing (Wichiensin, Bell, and Yang 2007).

In the context of the debate around the introduction of congestion pricing in New York City, Kheel (2008) recently proposed to completely eliminate the fare for public transportation by paying operating costs with the congestion charges. His own words, “[t]his more balanced plan will result in the equivalent of a \$20 after-tax pay raise for every transit-using worker in the city. Automobile drivers will benefit too, as traffic is vastly reduced” (p. 4), capture why rebates provide a more equitable solution than congestion pricing alone. Because having no fare may or may not be optimal from a system welfare perspective, we focus on finding the optimal level of subsidies. We assume that if congestion pricing is used, toll charges are already fixed. Specifically, we concentrate on mode decisions in the case of linear congestion costs and homogeneous demand. The main assumption of this article is that a city can set apart some funds that it will use to subsidize users of certain modes by offering a rebate on part of the fare. As Kheel said, rebates go to the population segment that selects the least-desirable modes, thus compensating users that switched out from their preferred choices. The fact that most public transportation systems in the world are subsidized provides anecdotal evidence that a mechanism based on subsidies is easier to accept by the constituents than congestion pricing. Well-chosen rebates lead to more efficient choices. Less people will drive, congestion will be reduced, and the total commute time will decrease. Eventually, some of the benefits will be transferred back to the provider of subsidies in the form of additional taxes, reduced CO₂ emissions, reduced health-care costs, etc. For example, operating expenses of companies that do deliveries will be reduced, thereby improving their bottom line. The additional taxes can be used to recover a fraction of the money that was set aside initially.

Cities do not have unlimited resources and, thus, cannot offer large rebates if they do not also implement congestion pricing. For that reason, we look at the problem of finding rebates that maximize user welfare, taking into account a limited budget. This budget relates to the value placed on the reduction of commute times. In the extreme case when commute times are all that matters and the budget is large, rebates will be set to make experienced costs equal to zero. (Compare this to the costs experienced by commuters under Kheel’s proposal, which are not

¹ Electronic toll collection systems—currently in use in cities that implement congestion pricing and in many tolled highways—eliminate the need to stop at a toll plaza. In general, these systems have three components: a toll tag, which is placed inside the vehicle; an overhead antenna, which reads the toll tag and collects the toll; and video cameras to identify toll evaders.

zero because commuters still face the disutility arising from the time invested to complete the trip.) On the other extreme, when the reduction in commute time is not deemed important or when the budget for rebates is small, rebates will not be offered and users will experience the full cost arising from the time and the fare or toll.

Even with optimal rebates, the coordination generated by this approach may not be enough to achieve a significant increase in welfare. Henceforth, we want to quantify the coordinating power of rebates. Koutsoupias and Papadimitriou (1999) defined the *price of anarchy* as the worst-case ratio of the social welfare under a user equilibrium attained without coordination to that with socially-optimal choices. This indicator has been used to estimate the potential increase in welfare provided by a given mechanism, and to gauge whether the opportunity cost is large enough to outweigh the implementation cost and justify its use. To answer this question, we compare the total welfare generated by optimal rebates to that when rebates are set to zero. We show that when the budget is large enough, one can have a transportation pattern that is significantly more efficient than the status quo.

Main Contributions and Structure of the Paper

Although others considered rebates implicitly (as negative taxes), to our knowledge, this is the first article that formally studies the computation of optimal rebates with the goal of coordinating a congestion game. Our main contribution is a mechanism that provides incentives for coordination that does not penalize participants, but instead rewards those that were worse-off without such a mechanism by offering them a rebate. Our social cost function explicitly considers the transfer payments to capture the cost of providing rebates, and the mechanism aims to minimize this more general expression of cost. Instead, most of the earlier articles that studied the coordinating power of tolls and taxes consider a social cost equal to the sum of costs for all participants, thus ignoring the costs and benefits of payments because they are transfers that stay in the system (see Beckmann, McGuire, and Winsten 1956; Bergendorff, Hearn, and Ramana 1997; Labbé, Marcotte, and Savard 1998 for classical references; Cole, Dodis, and Roughgarden 2006 is a notable exception that considers transfer payments as part of the social cost).

We consider a *Stackelberg game* in which the system owner (e.g., the city or the transportation authority) is the *leader* and the participants are *followers* (von Stackelberg 1934). In a first stage, the leader offers rebates in each arc; in a second stage, participants selfishly select arcs that have minimal cost, taking rebates into consideration. Focusing on the modal

choice problem, we characterize the optimal rebates in the case of affine cost functions and networks with multiple arcs that connect two nodes (the alternative modes of transportation are substitutes). Many examples of recent work in this area such as Engel, Fischer, and Galetovic (2004), Xiao, Yang, and Han (2007), Acemoglu and Ozdaglar (2007), and Wichiesin, Bell, and Yang (2007), also consider this type of simple networks. Although Labbé, Marcotte, and Savard (1998) present results for general networks, they do it for a simplified model that ignores congestion effects, which is an important feature of our model.

We first prove that if the system owner values the perceived cost more than rebates, then an optimal strategy for the leader is to refund each participant the perceived cost at each arc under a system optimal solution. When the system owner is more sensitive to the investment in rebates than to the perceived cost, it will offer rebates in the modes that are under-used. We also establish an upper bound on the proportion of participants that receive a positive rebate. Using our characterization of Stackelberg equilibria, we provide a polynomial-time algorithm that selects the arcs where rebates should be offered, and computes the optimal rebates for those arcs. This enables us to derive an explicit formula for the resulting social cost, from where we compute the price of anarchy, expressed as a function of the predisposition of the system owner to offer rebates. The main conclusion is that when the system owner is willing to offer rebates, the resulting solution has low social cost. Conversely, when the system owner cannot afford to provide significant rebates, the resulting outcome is close to a Wardrop equilibrium.

This paper is organized as follows. First, we review the literature in §2. In §3, we introduce the model and the performance measures of interest. Section 4 offers some results for general network topologies, while §5 focuses on instances with parallel arcs (substitutes) and characterizes the optimal rebates. In §6, we compute the price of anarchy for instances with affine cost functions. Finally, we offer some concluding remarks and open questions in §7.

2. Connections to the Literature

We work under the setting first described by Wardrop (1952). The corresponding equilibrium concept has been called a *Wardrop equilibrium*, which under mild conditions coincides with a Nash equilibrium (Haurie and Marcotte 1985). Although in some cases a system may be better off without a coordination mechanism because the overall implementation and operating costs may outweigh the potential benefits, equilibria have been found to be too inefficient in many applications of interest. This makes it necessary to

coordinate participants to mitigate the adverse effects of the misalignment of incentives. As imposing decisions to users is not an option in most real-world situations, equilibria can be improved by system (re)design (Roughgarden 2006), by considering routing part of the flow preemptively (Korilis, Lazar, and Orda 1995), or by using pricing mechanisms to create incentives (Bergendorff, Hearn, and Ramana 1997; Labbé, Marcotte, and Savard 1998). This article considers the third approach.

Even before the work of Vickrey (1955), economists such as Dupuit (1849), Pigou (1920), and Knight (1924) proposed to use pricing so participants internalize the externalities, defined as the additional cost they impose to others. If implemented properly, this results in equilibria that are efficient from a social welfare perspective. For a complete treatment of network pricing and many additional references, see, e.g., the book by Yang and Huang (2005).

We study a mechanism based on rebates. Rebates are used in logistics, supply chain management, and marketing, with the objective of revenue maximization as well as to create incentives for coordination (Gerstner and Hess 1991; Ali, Jolson, and Darmon 1994; Taylor 2002; Chen, Li, and Simchi-Levi 2007). We find the optimal rebates by solving a Stackelberg equilibrium problem, which structurally is a *mathematical program with equilibrium constraints* (MPEC). There are relatively standard optimization techniques to compute solutions to this type of problems. For a background on MPECs and solution methods, we refer the interested reader to the book by Luo, Pang, and Ralph (1996). One could get the optimal rebates and the corresponding modal choices from a Stackelberg equilibrium computed numerically; actually, computational studies are routinely used to analyze congestion-charging systems. In our case, though, finding the optimal rebates numerically is not enough for our purposes because such an analysis does not provide the structure needed to understand how much benefit the mechanism provides.

Recently, many authors have studied the maximum efficiency-loss under an equilibrium, using social welfare to measure the quality of solutions. Koutsoupias and Papadimitriou (1999) defined the price of anarchy as the largest possible ratio of the social cost at an equilibrium to the minimum attainable social cost (the term itself was coined by Papadimitriou 2001). Starting from the work of Roughgarden and Tardos (2002), the price of anarchy in transportation networks (the setting suggested by Wardrop 1952) has been characterized by Roughgarden (2003), Correa, Schulz, and Stier-Moses (2004), Chau and Sim (2003), and Perakis (2007), who successively considered more general assumptions. It turns out that equilibria of these

games are reasonably efficient; for example, when congestion costs increase linearly with flow, the extra total cost of an equilibrium does not exceed 33% more than that of a system optimum. For other typical classes of functions, the inefficiency is somewhat larger but bounded. Nevertheless, for practical purposes these inefficiencies are too high; even smaller improvements translate to big savings for societies and governments (recall the figures provided by the Urban Mobility Study). Hence, some researchers looked for improved measures of inefficiency (Friedman 2004; Qiu et al. 2006; Schulz and Stier-Moses 2006; Correa, Schulz, and Stier-Moses 2008), while others focused on mechanisms to improve the inefficiency itself. Some references that look at pricing mechanisms from the perspective of the price of anarchy are Koutsoupias (2004), Karakostas and Kolliopoulos (2005), Cole, Dodis, and Roughgarden (2006), Wichensin, Bell, and Yang (2007), Xiao, Yang, and Han (2007), and Yang, Xu, and Heydecker (2009).

The study of the inefficiency of equilibria has recently received increased attention from researchers in various communities such as Transportation, Operations Research, Operations Management, Economics, and Computer Science. Consequently, there is a growing amount of interdisciplinary literature on the price of anarchy. For example, some additional references in the application domains of telecommunication and distribution networks are the articles by Johari and Tsitsiklis (2004), Golany and Rothblum (2006), Perakis and Roels (2007), Acemoglu and Ozdaglar (2007), and Johari, Weintraub, and Van Roy (2009).

3. Description of the Model

In this section, we introduce the model and its necessary notation. We consider the framework of *network games*, originally introduced by Wardrop (1952) and first analyzed formally by Beckmann, McGuire, and Winsten (1956). An instance of our problem is given by a network, cost functions, a system owner, and participants. The network encodes the modal and route choices, and the cost functions associated to each arc model congestion and charges. The system owner defines the level of rebates, and participants—who are infinitesimally small—select a route from their origins to their destinations with minimum cost.

The network is represented by a directed graph (V, A) , where V is a set of vertices and A is a set of arcs. In general the graph may be arbitrary, but we will concentrate on the case where A is a set of parallel links that represent each of the modes. When possible, we will present results for general graphs to allow for route choice. For a total flow of x_i in an arc $i \in A$, the cost of traversing it is $c_i(x_i)$. Functions c_i , referred to as *cost functions*, are assumed to be affine on x_i for the

main results of this study. When possible, we will also consider more general cost functions that are nonnegative, nondecreasing, differentiable, and convex. Furthermore, we assume that cost functions are separable, meaning that the only argument of a cost function is the flow along that arc.

As we described in the introduction, the most typical example of this model is given by a urban network in which commuters have to decide between driving their cars, walking or taking one form of public transportation. Cost functions encode commute time, delays, and fares, all of which are assumed to be expressed in monetary units, and indicate the overall equivalent cost perceived by users for traversing a link. Although we do not explicitly model congestion pricing, it can be partially incorporated in our model by adding the corresponding charges to the cost functions.

The system owner offers rebates to elicit coordination. We denote the rebate for arc i by $s_i \geq 0$. As participants will not be reimbursed more than their cost, we restrict the actual reimbursement to not exceed $c_i(x_i)$. Hence, as the rebates are announced before participants make their selections, participants receive a rebate up to the cost of the arc. Indeed, the experienced cost is

$$c_i^s(x_i) := [c_i(x_i) - s_i]^+,$$

where $[y]^+$ denotes the positive part of y . Equivalently, the actual rebate equals $\min(s_i, c_i(x_i))$. Collectively, we denote the vector of all rebates with $s \in \mathbb{R}_+^A$.

Each participant selects the arc in A that corresponds to the mode of choice. For the results in which we also consider route selection, participants are associated with a pair of nodes, called an origin-destination pair (OD-pair), and have to select a path from their origins to their destinations. Let us denote the set of OD-pairs by K , the demand corresponding to OD-pair $k \in K$ by r_k , and the total demand $\sum_{k \in K} r_k$ by r . In addition, we refer to all the possible paths connecting an OD-pair $k \in K$ by \mathcal{P}_k and we let $\mathcal{P} := \bigcup_{k \in K} \mathcal{P}_k$. For the mode-choice model, there is a single OD pair that consists of the only two nodes.

We use flows to encode all participants' decisions, as specific identities are irrelevant. A flow x is feasible if it is nonnegative and it satisfies all demand constraints. Mathematically, this is represented by the set $\{x \in \mathbb{R}_+^{\mathcal{P}} : \sum_{P \in \mathcal{P}_k} x_P = r_k \text{ for all } k \in K\}$. The flow on an arc x_i is given by the sum over the paths $\sum_{P \in \mathcal{P} : P \ni i} x_P$.

Competition leads participants from the same OD-pair to select paths of cheapest equal cost because otherwise they would have an incentive to change their selection. This is the basis of the traditional solution concept called *Wardrop equilibrium* (Wardrop 1952).

DEFINITION 3.1. A flow x^{WE} is a Wardrop equilibrium of a network game (without rebates) if it is feasible, and for all k and all $P, Q \in \mathcal{P}_k$ such that $x_P^{\text{WE}} > 0$, $c_P(x^{\text{WE}}) \leq c_Q(x^{\text{WE}})$, where $c_P(x) := \sum_{i \in P} c_i(x_i)$.

The previous definition provides us with a solution concept that models the behavior of the second stage players:

DEFINITION 3.2. If the system owner selects the rebate vector s , participants select a solution x^s , which is a Wardrop equilibrium with respect to cost functions $[c_i(\cdot) - s_i]^+$.

For a given rebate vector s , the corresponding Wardrop equilibrium x^s always exists because the modified cost functions $[c_i(\cdot) - s_i]^+$ are continuous (Beckmann, McGuire, and Winsten 1956). In general, the equilibrium x^s need not be unique but if there are more than one, the prevailing costs under different equilibria are equal. Because any equilibrium can arise in practice, we consider an arbitrary one.

We now focus on the best strategy for the system owner. Because it is the leader of the Stackelberg game and it fixes the rebates knowing that participants are going to select a Wardrop equilibrium, its optimal strategy is to select the vector s that minimizes the *social cost*, defined as the sum of the costs of all parties in the game (Mas-Colell, Whinston, and Green 1995). This objective function includes the perceived cost experienced by each participant and the amount the system owner invests in rebates. As the system owner may be more sensitive to one of the terms than to the other, we consider a parameter $\rho \geq 0$ that transforms the rebate investment into social cost units. Section 6.1 provides further justification for this choice of social cost functions. (Note that we can alternatively define the social cost as the sum of the real costs that participants face by using a modified coefficient as shown in (1b).)

DEFINITION 3.3. The strategy (s, x^s) is a *Stackelberg equilibrium* if the vector of rebates s minimizes the *social cost*, defined as

$$C_\rho(s) := \underbrace{\sum_{i \in A} x_i^s [c_i(x_i^s) - s_i]^+}_{\text{participants' perceived cost}} + \rho \underbrace{\sum_{i \in A} x_i^s \min(c_i(x_i^s), s_i)}_{\text{cost of rebates}}, \quad (1a)$$

which can also be expressed as

$$\underbrace{\sum_{i \in A} x_i^s c_i(x_i^s)}_{\text{participants' real cost}} + (\rho - 1) \underbrace{\sum_{i \in A} x_i^s \min(c_i(x_i^s), s_i)}_{\text{cost of rebates}}. \quad (1b)$$

In this case, we refer to s as an *optimal rebate vector*.

The parameter ρ allows the system owner to control the tradeoff between the social cost of the solution and its investment. Alternatively, it can be viewed as the Lagrangian multiplier of the system owner's budget constraint. In fact, $1/\rho$ represents the investment

the system owner is willing to commit to make the participants' perceived cost decrease by one unit:

- $\rho = 1$ corresponds to the situation in which the system owner is only interested in minimizing the participants' real cost $\sum_{i \in A} x_i c_i(x_i)$, regardless of the rebate cost (see (1b)).
- $\rho = +\infty$ corresponds to the situation in which the system owner does not want to spend any money on rebates. Here, the outcome will be a Wardrop equilibrium, as without rebates.
- Values of $\rho < 1$ correspond to the case where the network planner values the participants' perceived cost more than its own investments.

As we said in §2, the Stackelberg equilibrium can be found by solving an MPEC. If the leader wants to compute optimal rebates for a particular instance, there are relatively standard optimization techniques to solve this problem, even if more constraints are added to the problem (e.g., restrict rebates to a subset of arcs, or impose that rebates may not exceed monetary charges such as tolls or fares). Instead, we will work with the optimality conditions of this problem to explicitly characterize the Stackelberg equilibrium. This will allow us to design an efficient algorithm and to find the worst-case inefficiency of the corresponding equilibrium.

Not only do we want to compare the social cost of different solutions with rebates, but we also want to compare using rebates to not using them. Therefore, another measure of interest is the participants' real cost, represented by the objective function $C(x) := \sum_{i \in A} x_i c_i(x_i)$. The following definition captures the situation when the system owner controls the whole system.

DEFINITION 3.4. A flow x^{SO} is a system optimum if it is feasible and minimizes $C(\cdot)$.

The following proposition draws on the first-order optimality conditions to the mathematical program that defines a system optimum.

PROPOSITION 3.5 (BECKMANN, MCGUIRE, AND WINSTEN 1956). *For instances with differentiable and convex cost functions, a flow x^{SO} is a system optimum if and only if it is a Wardrop equilibrium with respect to the modified cost functions $c_i^*(x_i) := c_i(x_i) + x_i c'_i(x_i)$, where $c'_i(x)$ is the derivative of $c_i(x)$ with respect to x .*

Note that if $\rho \geq 1$, the social cost of a Stackelberg equilibrium (s, x^s) satisfies

$$C(x^{SO}) \leq C_\rho(s) \leq C(x^{WE}). \tag{2}$$

The lower bound follows from (1b) because its second term is nonnegative, and the upper bound comes from the feasibility of $s = 0$ because $C(x^{WE}) = C_\rho(0)$.

3.1. Examples

In this section we introduce two concrete instances that will be the running examples for the rest of the article. These instances will be used to illustrate the different concepts and calculations along the way.

INSTANCE 1 (ROUGHGARDEN AND TARDOS 2002). The first instance represents a competitive situation first described by Pigou (1920). As illustrated in Figure 1, participants must select one of two available modes: the first is expensive but its cost is not influenced by demand, while the second one is cheap under low demand but becomes expensive if it attracts many participants. This instance models a decision that commuters make daily in many cities. A person can use mass transit and experience an almost constant but large commute time, or can drive to (hopefully) experience a short commute while being exposed to the possibility of congestion.

The total demand in this instance is equal to 1, composed of an infinite number of price-taking users. The Wardrop equilibrium routes all flow in the lower arc because all participants take lowest-cost routes. Under this solution, $C(x^{WE}) = 1$. To exploit the effects of congestion, the system optimum assigns half of the participants to each mode, implying that $C(x^{SO}) = 3/4$.

If $\rho \leq 1$, the system owner will propose rebates equal to $(1, 1/2)$, which is the vector of prevailing costs under the system optimum. This results in an equilibrium that matches the system optimum. Actually, §4.1 shows that, for arbitrary instances, rebates lead to the system optimum when $\rho \leq 1$ because experienced costs are zero. Let us now consider the case $\rho > 1$. It does not make sense to offer a rebate in both arcs because subtracting a constant everywhere will not change the equilibrium. Therefore, the system owner should only consider giving a rebate in the upper arc (the lower one is always cheaper so it should not be subsidized). Denoting this rebate by $s \in [0, 1]$, the perceived cost on this arc equals $1 - s$. Therefore, the corresponding Wardrop equilibrium x^s is the flow that routes s units in the upper arc. After some algebra, $C_\rho(s) = 1 - s + \rho s^2$. The minimum, which provides the Stackelberg equilibrium, is $s = 1/(2\rho)$ and achieves a social cost of $1 - 1/(4\rho)$.

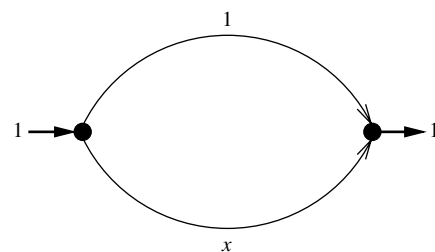


Figure 1 Pigou's Example
 Note. Arcs are labeled with their cost functions.

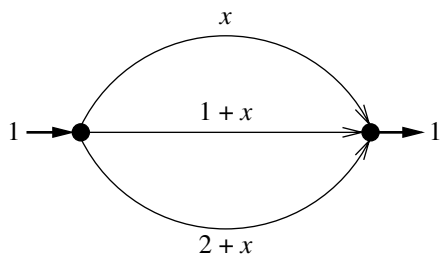


Figure 2 Running Example

Note. Arcs are labeled with their cost functions.

INSTANCE 2. The second network is similar to Pigou’s but contains an extra mode. As depicted in Figure 2, the three modes, numbered from 1 to 3 for simplicity, have cost functions equal to $c_i(x_i) := (i - 1) + x_i$. At the Wardrop equilibrium, all participants select the first mode, and therefore $C(x^{\text{WE}}) = 1$. The system optimum is given by the flow $(3/4, 1/4, 0)$, with total cost $C(x^{\text{SO}}) = 7/8$. Finally, an optimal rebate vector for $\rho > 1$ is $s = (0, 1/(2\rho), y)$, with $0 \leq y \leq 1 + 1/(4\rho)$. The corresponding Wardrop equilibrium x^s is $(1 - 1/(4\rho), 1/(4\rho), 0)$, and its social cost equals $C_\rho(s) = 1 - 1/(8\rho)$.

3.2. An Application to Logistics

The framework that we consider can readily be used to model competition in other settings such as telecommunication and distribution networks. This section briefly comments on an application in the area of logistics.

We consider a freight company that sends goods across a network. The system owner models the corporate headquarters while participants model business units that manage different markets. The system is not controlled centrally; units make their own decisions about how goods are transshipped across the network, considering their individual costs. This network is composed of resources, which may represent different carriers that transport freight or facilities that process it. Resources include sorting facilities, warehouses, flight legs, airports, ship routes, ports, canals, etc. Some of those resources belong to the unit, others belong to the company and are shared between units, and some are controlled by third parties. Resources that are not controlled directly by the unit will be priced according to the laws of offer and demand. Hence, competition for a resource will drive its price up, which can be represented by cost functions (in this case cost-demand curves). We assume that units are not big enough to influence prices independently (i.e., they are price-taking).

Units select a set of resources to transship their goods at minimum cost, and are rewarded by the

profits they generate.² Cost-demand curves create externalities between units, which is what causes competition among them. If nothing is done, the stable situation would be an equilibrium among the business units that is generally inefficient in terms of the company’s total profit. Realizing the problem, the company can compute the system optimum ignoring the goals of the individual business units, and find the rebates that it should offer for each resource. In this way, the headquarters will be offering incentives that help align business units into maximizing the company’s profits.

4. General Network Topologies

We start our study of the structural characteristics of Stackelberg equilibria. In this section, we consider general network topologies, with possibly several OD-pairs. We start by considering the case of the system owner assigning more value to the participants’ perceived cost than to its own rebate investment, and characterize the optimal strategy when setting the rebates. Later, we turn into the opposite case and provide some properties that will be used to characterize optimal rebates.

4.1. Small ρ

This section focuses on achieving a fully efficient coordinated solution for the case of $\rho \leq 1$ and networks with arbitrary topology. As suggested in §3.1, let us consider the rebate vector given by $s_i = c_i(x_i^{\text{SO}})$ for all $i \in A$. With those rebates, the system optimum x^{SO} is an equilibrium for the participants’ game because participants experience a cost equal to zero (which is the absolute minimum because of the nonnegativity of modified cost functions).

Beckmann, McGuire, and Winsten (1956) proved that payments equal to marginal costs at the system optimum also lead to a system optimum (see Proposition 3.5); recalling that $c_i^*(x_i) = c_i(x_i) + x_i c'_i(x_i)$, this corresponds to negative rebates $s_i = -x_i^{\text{SO}} c'_i(x_i^{\text{SO}})$. Moreover, any convex combination of optimal transfers payments (tolls or rebates) is also optimal (Bergendorff, Hearn, and Ramana 1997), which implies that the set of transfers payments that lead to system optimality is a polyhedron. We summarize these claims in the following remark.

² All of the results valid for arbitrary networks are also valid in the more general setting of *nonatomic congestion games* (Rosenthal 1973). In this case, business units will select one set of resources from a list of feasible sets, without insisting that these sets have to be paths. This more general competitive situation is called nonatomic because participants are price taking, and a congestion game because participants are anonymous and costs of resources depend only on the number of participants selecting them.

REMARK 4.1. When rebates equal

$$s = [(1 - \kappa)c_i(x^{\text{SO}}) - \kappa(x_i^{\text{SO}}c'_i(x_i^{\text{SO}}))]_{i \in A}$$

with $0 \leq \kappa \leq 1$, a system optimal solution x^{SO} is at equilibrium. Here, positive values of s_i represent rebates and negative values represent payments. Moreover, if cost functions are strictly increasing, the corresponding equilibrium x^s is unique.

The next proposition shows that the previously-mentioned rebates are optimal when cost functions are strictly increasing. It turns out that this is the only optimal vector and leads to a unique second-stage equilibrium, which matches the system optimum. If we only consider weakly increasing functions, then a system optimum is always at equilibrium for that rebate vector but there may be other equilibria. In that case, though, an optimal rebate vector may not exist.

PROPOSITION 4.2. Assume that $\rho \leq 1$ and that cost functions are strictly increasing. For arbitrary networks, a Stackelberg equilibrium (s, x^s) satisfies that $s = (c_i(x_i^{\text{SO}}))_{i \in A}$ and $x^s = x^{\text{SO}}$. This equilibrium achieves a social cost of $C_\rho(s) = \rho C(x^{\text{SO}})$.

PROOF. Considering s as described in the proposition, let us prove that x^s has to be equal to x^{SO} . The Stackelberg flow x^s is a Wardrop equilibrium under the modified cost functions. The variational inequality characterization of Wardrop equilibria (Smith 1979) and the choice of s implies that for a feasible flow x , $\sum_{i \in A} (x_i - x_i^{\text{SO}})[c_i(x_i^s) - c_i(x_i^{\text{SO}})]^+ \geq 0$. Because the optimal flow x^{SO} is feasible, we have $\sum_{i \in A} (x_i^{\text{SO}} - x_i) \cdot [c_i(x_i^s) - c_i(x_i^{\text{SO}})]^+ \geq 0$. The summands vanish on arcs i such that $x_i^s \leq x_i^{\text{SO}}$, and are strictly negative on arcs i for which $x_i^s > x_i^{\text{SO}}$. Consequently, $x_i^s \leq x_i^{\text{SO}}$ for all $i \in A$, resulting in $x^s = x^{\text{SO}}$ because x^{SO} is a feasible flow without cycles (because it minimizes the participants' real cost and link cost functions are strictly increasing). Evaluating the social cost, we compute that $C_\rho(s) = \rho C(x^{\text{SO}})$.

Let us now show that this choice of s provides the same social cost as an optimal rebate vector s^* . Using the nonnegativity of the first term of (1a) and the feasibility of s , respectively, $\rho \sum_{i \in A} x_i^{s^*} \min(c_i(x_i^{s^*}), s_i^*) \leq C_\rho(s^*) \leq C_\rho(s)$, from where $\sum_{i \in A} x_i^{s^*} \min(c_i(x_i^{s^*}), s_i^*) \leq C(x^{\text{SO}})$. Bounding each of the terms in (1b) separately, $C_\rho(s^*) \geq C(x^{\text{SO}}) + (\rho - 1)C(x^{\text{SO}}) = \rho C(x^{\text{SO}})$, where we used that x^{SO} minimizes $C(\cdot)$ and that $\rho \leq 1$. Hence, $\rho C(x^{\text{SO}})$ is a lower bound for the optimal social objective that is attained at s , which establishes the proposition. \square

4.2. Large ρ

In this section we consider that $\rho > 1$. For constant cost functions, it is optimal to offer no rebates. Indeed, when $s = 0$, the participants' real cost under a Nash equilibrium equals that of a system optimum and the

cost of rebates is zero. Because both terms of (1b) equal a lower bound, this choice of s is optimal for the leader. Rebates are useful only in the presence of congestion. (Note that we get to a similar conclusion in the model of Labbé, Marcotte, and Savard 1998, who assumed that there is no congestion and that the leader is a revenue-maximizer.)

We will characterize the benefits of offering rebates by studying the structure of Stackelberg equilibria. We start by proving that under an optimal rebate vector there is always at least one used arc with positive experienced cost, and one used arc in which no rebate is offered. We let $\mathcal{F} \subseteq A$ be the set of arcs with positive flow under the equilibrium, which we partition into sets \mathcal{F}_s , containing arcs with positive rebates, and \mathcal{F}_0 , containing arcs with no rebates.

DEFINITION 4.3. For a given rebate vector s , define $\mathcal{F} := \{i \in A \mid x_i^s > 0\} = \mathcal{F}_s \cup \mathcal{F}_0$, where $\mathcal{F}_s := \{i \in \mathcal{F} \mid s_i > 0\}$ and $\mathcal{F}_0 := \{i \in \mathcal{F} \mid s_i = 0\}$.

Without loss of generality, we will sometimes assume that rebates for arcs in $A \setminus \mathcal{F}$ are zero. Indeed, if an unused arc has a positive rebate, it will still be unused without the rebate. Consequently, the corresponding Wardrop equilibrium and all the aggregate measures we considered do not change when the rebate is removed. For example, for the Stackelberg equilibrium of Instance 2, we have that $\mathcal{F}_0 = \{1\}$ and $\mathcal{F}_s = \{2\}$. The third arc does not belong to \mathcal{F} because its flow is zero.

LEMMA 4.4. Assume that $\rho > 1$ and that all cost functions are strictly increasing. For an arbitrary network, if (s, x^s) is a Stackelberg equilibrium, then there exists an arc $i \in \mathcal{F}$ such that $s_i < c_i(x_i^s)$.

PROOF. Assume that all perceived costs are zero, i.e., $s_i \geq c_i(x_i^s)$ for all $i \in \mathcal{F}$. Without loss of generality, it is enough that $s_i = c_i(x_i^s)$ for all those arcs. Then the social cost equals $C_\rho(s) = \rho C(x^s) \geq \rho C(x^{\text{SO}})$. As stated in §4.1, the social cost $\rho C(x^{\text{SO}})$ can be attained with rebates $(c_i(x_i^{\text{SO}}))_{i \in A}$. Since s was assumed to be an optimal rebate vector, we must have that $C(x^s) = C(x^{\text{SO}})$, from where we see that x^s is a system optimum. Because of Proposition 3.5, x^s is at equilibrium with respect to modified costs $c_i(x_i^s) + x_i^s c'_i(x_i^s)$.

As perceived costs are zero and cost functions are strictly increasing, $s_i > 0$ for all $i \in \mathcal{F}$, or equivalently $\mathcal{F}_0 = \emptyset$. Hence, there exists a small enough $\epsilon > 0$ such that $\tilde{s} \geq 0$, where

$$\tilde{s}_i := \begin{cases} s_i - \epsilon(c_i(x_i^s) + x_i^s c'_i(x_i^s)) & i \in \mathcal{F}, \\ 0 & i \in A \setminus \mathcal{F}. \end{cases}$$

Under rebates \tilde{s} and flow x^s , the perceived cost on each arc is $[c_i(x_i^s) - \tilde{s}_i]^+ = \epsilon(c_i(x_i^s) + x_i^s c'_i(x_i^s))$ for $i \in \mathcal{F}$. Similarly, $[c_i(x_i^s) - \tilde{s}_i]^+ = c_i(x_i^s) \geq \epsilon c_i(x_i^s) = \epsilon(c_i(x_i^s) + x_i^s c'_i(x_i^s))$ for $i \in A \setminus \mathcal{F}$. The last two equations

imply that x^s is at equilibrium under rebates \tilde{s} , and the perceived cost on each used arc is strictly positive. Finally, x^s is the unique equilibrium under \tilde{s} since the potential function $F(x) := \sum_{i \in A} \int_0^{x_i} [c_i(z_i) - \tilde{s}_i]^+ dz_i$ is convex in general, strictly convex in a vicinity of x^s as the cost functions are strictly increasing, and achieves a minimum at x^s . (We refer the reader to Beckmann, McGuire, and Winsten 1956 for details on the characterization of Wardrop equilibria with this type of potential function.) Consequently,

$$\begin{aligned} C_\rho(\tilde{s}) &= \sum_{i \in A} x_i^s c_i(x_i^s) + (\rho - 1) x_i^s \tilde{s}_i \\ &= C_\rho(s) - \epsilon(\rho - 1) \sum_{i \in \mathcal{J}} x_i^s \underbrace{(c_i(x_i^s) + x_i^s c'_i(x_i^s))}_{>0} < C_\rho(s), \end{aligned}$$

which is a contradiction to the optimality of s . \square

When we presented the examples in §3.1, we mentioned that it cannot be optimal to offer rebates in all arcs. The next lemma generalizes this observation to any network topology. It shows that, if all arcs are used, then \mathcal{J}_0 is necessarily nonempty. In §5, we will further generalize this lemma to instances in which not all arcs are used, but under the restriction that the network has parallel links. Notice that in the case of a general network without the restriction that all arcs are used, we do not know if \mathcal{J}_0 could be empty. If such generalization were valid, Lemma 4.4 would not be necessary because it would be implied by this result. Indeed, because cost functions are nonnegative and strictly increasing, any arc in \mathcal{J}_0 would experience a positive cost because it is used and has no rebate.

LEMMA 4.5. *Assume that $\rho > 1$ and that all cost functions are strictly increasing. For an arbitrary network, if (s, x^s) is a Stackelberg equilibrium and all arcs are used, then there exists an arc $i \in \mathcal{J}$ such that $s_i = 0$.*

PROOF. With the purpose of deriving a contradiction, let us assume that s is an optimal vector of rebates such that $s_i > 0$ for all $i \in \mathcal{J}$. We will show that we can decrease the rebates while maintaining the same user equilibrium. Note that, unless the network only consists of parallel links, subtracting a constant from all rebates may change the user equilibrium because it would make longer paths more attractive to users. Instead, the proposed rebates are such that the resulting perceived cost on all links is a multiple of the original perceived costs. Let us therefore consider new rebates $\tilde{s} = [c_i(x_i^s) - \eta[c_i(x_i^s) - s_i]^+]_{i \in A}$, where

$$\eta := \min_{i \in \mathcal{J}} \frac{c_i(x_i^s)}{[c_i(x_i^s) - s_i]^+}.$$

The definition implies that $\tilde{s} \geq 0$ and Lemma 4.4 implies that $\eta < \infty$, so the new rebates are well-defined. The perceived cost for arc i under the new

rebates equals $[c_i(x_i^s) - \tilde{s}_i]^+ = \eta[c_i(x_i^s) - s_i]^+$, meaning that x^s is also at equilibrium under \tilde{s} . Furthermore, as $s > 0$, we have that $\eta > 1$ and $\tilde{s} \leq s$. Hence, looking at (1b), the participants' real cost is unchanged, whereas the cost of rebates strictly decreases because $\tilde{s}_i = 0$ for the argument i achieving the minimum. This contradicts the optimality of s . \square

5. Networks with Parallel Links

Equipped with the structural results of the previous section, we now embark in the design of an efficient algorithm for computing Stackelberg equilibria. The outline of the procedure described in this section is as follows. First, we will partition arcs into those in which rebates must be offered, those in which no rebates must be offered and those that are not used in an equilibrium. With this partition, we will compute the actual rebates for the corresponding arcs.

We focus on networks in which participants have to select exactly one out of many possible arcs. This primarily models the mode choice problem but one can also use it for other applications in which users choose among substitutes. The network topology that corresponds to this situation comprises two nodes joined by several parallel arcs (see Figure 3). Networks with parallel arcs extend the classic two-route network introduced by Pigou (1920). They have been widely used because of their relevance to practical applications—such as transportation, telecommunication, scheduling, and resource allocation problems—and because of their tractability (see, e.g., Korilis, Lazar, and Orda 1995; Koutsoupias and Papadimitriou 1999; Roughgarden 2004; Engel, Fischer, and Galetovic 2004; de Palma and Picard 2006; de Palma, Kilani, and Lindsey 2007; Wichiansin, Bell, and Yang 2007; Xiao, Yang, and Han 2007; Acemoglu and Ozdaglar 2007; Johari, Weintraub, and Van Roy 2009). Note that the restriction to simple topologies seems necessary if we hope to find the optimal rebates in polynomial time because Cole, Dodis, and Roughgarden (2006) proved that finding

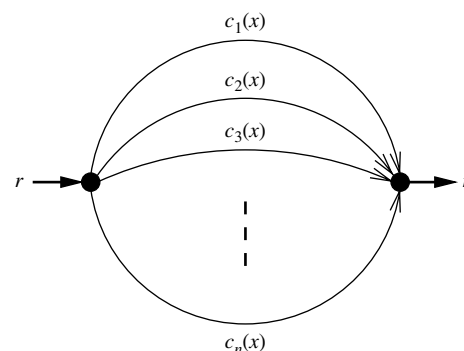


Figure 3 A Network with Parallel Arcs

optimal taxes in general networks with affine cost functions is hard.³ Finally, we only consider the case of $\rho > 1$, because the optimal rebates for $\rho \leq 1$ were already found in §4.1. We can assume without loss of generality that $s_i \leq c_i(x_i^s)$, as it is never beneficial to offer more.

Consider a Stackelberg equilibrium (s, x^s) of an instance in which cost functions are strictly increasing. The equilibrium conditions imply that there is a constant $L_\rho \geq 0$ such that

$$L_\rho = c_i(x_i^s) - s_i \quad \forall i \in \mathcal{F}, \quad (3a)$$

$$L_\rho \leq c_i(0) - s_i \quad \forall i \in A \setminus \mathcal{F}. \quad (3b)$$

Moreover, Lemma 4.4 implies that L_ρ has to be strictly positive. Hence, $c_i(x_i^s) > s_i$ for all $i \in A$. For networks with parallel arcs, then, we do not need to enforce the constraint that the system owner cannot offer rebates that are larger than the cost of arcs. In this case (1) simplifies to $C_\rho(s) = \sum_{i \in A} x_i^s [c_i(x_i^s) + (\rho - 1)s_i]$.

REMARK 5.1. The positivity of L_ρ also implies that when cost functions are strictly increasing there is a unique Wardrop equilibrium corresponding to the optimal s because the potential function $F(x) = \sum_{i \in A} \int_0^{x_i} [c_i(z) - s_i]^+ dz$ is strictly convex in a vicinity of x^s . Later, we shall prove that in this case the optimal s is also unique.

Going back to the examples of §3.1, it is not hard to check that L_ρ for Instances 1 and 2 equals $1 - 1/(2\rho)$ and $1 - 1/(4\rho)$, respectively.

5.1. General Cost Functions

We start with general cost functions and then, in the next section, switch to the particular case of affine cost functions. This section proves a result that will allow us to decide for which arcs we must offer positive rebates. To get there, we first have to present a series of lemmas. The first one establishes that a rebate vector that is optimal for a given network is also optimal when some unused arcs are taken out. In other words, removing $i \in A \setminus \mathcal{F}$ does not affect the optimality of s . Missing proofs are given in the appendix.

LEMMA 5.2. *Consider a network with parallel arcs and an optimal rebate vector s . If l is an arc in $A \setminus \mathcal{F}$, then the vector s with the entry corresponding to l removed is optimal for a similar instance with arc l removed.*

³ Cole, Dodis, and Roughgarden (2006, Theorem 6.2) prove that an approximation algorithm with guarantee better than $4/3 - \epsilon$ cannot exist unless $P = NP$. Although their reduction does not work for our problem, we conjecture that finding the optimal rebates in a general network with affine cost functions is also NP-hard because of the similarity between their social cost function and (1b) (see also §6.1). Another evidence in this direction is given by Labbé, Marcotte, and Savard (1998), who prove that computing taxes and rebates that maximize the leader's profit is an NP-hard problem, even when the network is not subject to congestion effects.

Notice that the previous lemma generalizes Lemma 4.5 to an arbitrary instance with parallel arcs. Indeed, Lemma 5.2 implies that an optimal rebate vector s is still optimal for the network consisting only of arcs in \mathcal{F} . Because that instance makes use of all arcs, it must contain at least one arc without rebate.

In the following propositions, we derive necessary conditions for a rebate vector s to be optimal. The next proposition shows that the optimal rebates satisfy the following equilibrium conditions: rebates are offered only in arcs for which the expression $c_i^*(\cdot)$ is minimal. This is implied by the first-order optimality conditions of the MPEC that characterizes the optimal rebates. Contrast this to Proposition 3.5 that states that in a system optimum, *participants* are assigned only to arcs for which the expression $c_i^*(\cdot)$ is minimal.

PROPOSITION 5.3. *Consider a network with parallel arcs and strictly increasing and differentiable cost functions, and let (s, x^s) be a Stackelberg equilibrium. There exists $V_\rho > 0$ such that*

$$V_\rho = c_i(x_i^s) + x_i^s c'_i(x_i^s) \quad \forall i \in \mathcal{F}_s, \quad (4a)$$

$$V_\rho \leq c_i(x_i^s) + x_i^s c'_i(x_i^s) \quad \forall i \in A \setminus \mathcal{F}_s. \quad (4b)$$

From (3a) and (4), we get that there exists a constant $D_\rho := 2L_\rho - V_\rho$ such that

$$D_\rho = c_i(x_i^s) - x_i^s c'_i(x_i^s) - 2s_i \quad \forall i \in \mathcal{F}_s, \quad (5a)$$

$$D_\rho \geq c_i(x_i^s) - x_i^s c'_i(x_i^s) - 2s_i \quad \forall i \in \mathcal{F}_0. \quad (5b)$$

The common perceived cost at equilibrium therefore equals $L_\rho = (V_\rho + D_\rho)/2$. Comparing the expressions, it is clear that $D_\rho < L_\rho < V_\rho$. For example, looking at a Stackelberg equilibrium of Instance 2, the constants are $V_\rho = 1 + 1/(2\rho)$ and $D_\rho = 1 - 1/\rho$.

In the sequel, we will make extensive use of the following definition to characterize and to compute optimal rebates:

DEFINITION 5.4. For $X \subseteq A$, let $K(X) := \sum_{i \in X} c'_i(x_i^s)^{-1}$. For the special case of an empty set, it is assumed that $K(\emptyset) := 0$.

The following technical lemma provides a formula that will be useful later. Its proof considers another feasible direction from the optimal rebate vector.

LEMMA 5.5. *Consider a network with parallel arcs and strictly increasing and differentiable cost functions. Letting (s, x^s) be a Stackelberg equilibrium, then*

$$\sum_{i \in \mathcal{F}_s} \left(x_i^s K(\mathcal{F}) + \frac{s_i}{c'_i(x_i^s)} K(\mathcal{F}_0) \right) = \frac{r}{\rho} K(\mathcal{F}_s). \quad (6)$$

Using the previous results, we can characterize the sets \mathcal{F}_0 and \mathcal{F}_s , which will allow us to compute the optimal rebates.

PROPOSITION 5.6. *Consider a network with parallel arcs and strictly increasing and differentiable cost functions. Letting (s, x^s) be a Stackelberg equilibrium, for all $i \in A$ we have that*

$$i \in \mathcal{F}_0 \Leftrightarrow D_\rho \geq c_i(x_i^s) - x_i^s c'_i(x_i^s), \quad (7a)$$

$$i \in A \setminus \mathcal{F} \Leftrightarrow V_\rho \leq c_i(0). \quad (7b)$$

PROOF. We start with (7a). The forward implication is (5b). Conversely, consider $i \in A$, and assume that $D_\rho \geq c_i(x_i^s) - x_i^s c'_i(x_i^s)$. If $i \in A \setminus \mathcal{F}$, then $x_i^s = 0$ and $c_i(0) \leq D_\rho < L_\rho$, contradicting the Wardrop equilibrium condition. If $i \in \mathcal{F}_s$, then (5a) implies that $c_i(x_i^s) - x_i^s c'_i(x_i^s) = D_\rho + 2s_i > D_\rho$, yielding a contradiction again.

The forward implication of (7b) follows from (4b). Conversely, consider an $i \in A$, and assume that $V_\rho \leq c_i(0)$. If $i \in \mathcal{F}_0$, then $c_i(0) < L_\rho < V_\rho$, which yields a contradiction. If $i \in \mathcal{F}_s$ then (4a) implies that $c_i(0) < c_i(x_i^s) + x_i^s c'_i(x_i^s) = V_\rho$, which is again a contradiction. \square

In other words, we have the following partition of the arcs according to the expression $c_i(x_i^s) - x_i^s c'_i(x_i^s)$: considering $i_0 \in \mathcal{F}_0$, $i_s \in \mathcal{F}_s$, and $j \in A \setminus \mathcal{F}$, we have

$$\begin{aligned} c_{i_0}(x_{i_0}^s) - x_{i_0}^s c'_{i_0}(x_{i_0}^s) &\leq D_\rho < c_{i_s}(x_{i_s}^s) - x_{i_s}^s c'_{i_s}(x_{i_s}^s) \\ &< V_\rho \leq c_j(0). \end{aligned} \quad (8)$$

This characterizes which arcs are used naturally because they are cheap, which arcs are used because of the rebates offered, and which arcs are not used, even having the possibility of offering rebates, because they are too expensive. Of course, to use this result constructively one would first need to know the Stackelberg equilibrium. In the next section, we will see how to work around that problem for affine cost functions. Going back to Instance 2, one can see that $c_1(x_1^s) - x_1^s c'_1(x_1^s) = 0 \leq D_\rho = 1 - 1/\rho < c_2(x_2^s) - x_2^s c'_2(x_2^s) = 1 < V_\rho = 1 + 1/(2\rho) \leq c_3(0) = 2$.

We can now use the, so far partial, characterization of Stackelberg equilibria to determine how many participants extract a benefit from the availability of rebates in the network.

PROPOSITION 5.7. *Consider a network with parallel arcs and strictly increasing and differentiable cost functions, and let (s, x^s) be a Stackelberg equilibrium. The proportion of participants that receive a rebate is strictly lower than $1/\rho$.*

PROOF. Assume that $\mathcal{F}_s \neq \emptyset$ because otherwise the claim is obvious. Dividing (6) by $K(\mathcal{F}_s)$,

$$\begin{aligned} \frac{r}{\rho} &= \frac{K(\mathcal{F}_0)}{K(\mathcal{F}_s)} \sum_{i \in \mathcal{F}_s} \left(x_i^s + \frac{s_i}{c'_i(x_i^s)} \right) + \sum_{i \in \mathcal{F}_s} x_i^s \\ &= K(\mathcal{F}_0)(V_\rho - L_\rho) + \sum_{i \in \mathcal{F}_s} x_i^s. \end{aligned}$$

Therefore, $\sum_{i \in \mathcal{F}_s} x_i^s / r = 1/\rho - K(\mathcal{F}_0)(V_\rho - L_\rho) / r < 1/\rho$, as we wanted to show. \square

As expected, there is a strong correlation between how many participants respond to the incentive and the gains in the social cost that arise from it. The previous bound turns out to be tight as demonstrated by the following instance.

INSTANCE 3. Consider a network similar to that depicted in Figure 1 but with cost functions $c_1(x) = 1 - (1 - \epsilon)/\rho + \alpha x$ and $c_2(x) = x$, where $0 < \epsilon < 1$ and $\alpha > 0$. Using results we will develop in §5.2, we must have that $\mathcal{F}_0 = \{2\}$ and $\mathcal{F}_s = \{1\}$ (because $b_2 < L_\infty(1 - 1/\rho) = 1 - 1/\rho < b_1 < L_\infty(1 + 1/\rho)$; see the next section for the notation). Hence, the rebate $s = (\epsilon/(2\rho), 0)$ is optimal and the corresponding equilibrium is given by

$$x^s = \left(\frac{2 - \epsilon}{2(1 + \alpha)\rho}, 1 - \frac{2 - \epsilon}{2(1 + \alpha)\rho} \right).$$

The proportion of participants that receive positive rebates is x_1^s , which tends to $1/\rho$ as ϵ and α tend to 0.

5.2. Affine Cost Functions

Having derived properties for general cost functions, this section considers instances with affine cost functions and explicitly provides expressions for the optimal rebates. Instances with this type of cost functions are rich enough for many congestion phenomena to appear. For example, the well-known Braess paradox was initially formulated with affine cost functions (Braess 1968). Even for applications in which cost functions are more complex, an affine approximation can already show evidence of first-order effects (Acemoglu and Ozdaglar 2007; Johari, Weintraub, and Van Roy 2009). We denote the cost function on arc $i \in A$ by $c_i(x) = a_i x + b_i$, with $a_i > 0$ and $b_i \geq 0$. Without loss of generality, we consider that arcs are sorted according to b_i , so we have that $b_1 \leq b_2 \leq \dots \leq b_{|A|}$. For ease of notation, we let $[i] := \{1, \dots, i\}$, and $b_{|A|+1} = +\infty$.

In the case of affine functions, we can simplify some of the formulas we provided in previous sections. For example, Definition 5.4 becomes $K(X) = \sum_{i \in X} 1/a_i$ for $X \subset A$. Notice also that a consequence of (5) is that $D_\rho \geq 0$ and $s_i \leq b_i/2$ for all $i \in \mathcal{F}_s$. Furthermore, (8) allows us to partition the arcs into the sets \mathcal{F}_0 , \mathcal{F}_s , and $A \setminus \mathcal{F}$ as follows.

PROPOSITION 5.8. *Consider a network with parallel arcs and affine cost functions. If we consider $i_0 \in \mathcal{F}_0$, $i_s \in \mathcal{F}_s$, and $j \in A \setminus \mathcal{F}$, then $b_{i_0} \leq D_\rho < b_{i_s} < V_\rho \leq b_j$.*

The following lemma and theorem show that if we know how the arcs are partitioned, we can compute the optimal rebate values for all arcs.

LEMMA 5.9. Consider a network with parallel arcs and affine cost functions, and let (s, x^s) be a Stackelberg equilibrium. If rebates are beneficial (i.e., if $\mathcal{F}_s \neq \emptyset$), then

$$D_\rho = \frac{1}{K(\mathcal{F}_0)} \left(r \frac{\rho - 1}{\rho} + \sum_{i \in \mathcal{F}_0} \frac{b_i}{a_i} \right),$$

$$V_\rho = \frac{1}{K(\mathcal{F})} \left(r \frac{\rho + 1}{\rho} + \sum_{i \in \mathcal{F}} \frac{b_i}{a_i} \right).$$

If \mathcal{F}_0 is known, making use of the previous lemma, we can compute the optimal rebates using the relations that we developed in the previous section. This result implies that, essentially, there is a unique optimal rebate vector.

THEOREM 5.10. Consider a network with parallel arcs and affine cost functions. Then, the optimal rebates must satisfy that

$$s_i = \left[\frac{b_i - D_\rho}{2} \right]^+$$

for all $i \in \mathcal{F}$. Moreover, if this formula is used for all arcs, the corresponding solution (s, x^s) is a Stackelberg equilibrium.

PROOF. Consider an arc $i \in \mathcal{F}$. If $i \in \mathcal{F}_0$, then $s_i = 0$ by definition and this agrees with the proposed formula because of Proposition 5.8. If $i \in \mathcal{F}_s$, then solving for s_i in (5a) also gives the proposed formula.

Now consider using the proposed formula for all $i \in A$. We must prove that each arc $j \in A \setminus \mathcal{F}$ is not used under the corresponding Wardrop equilibrium. Proposition 5.8 implies that $b_j > V_\rho$. Therefore, the rebate computed by the theorem is positive and $b_j - s_j = (b_j + D_\rho)/2$. We conclude that the experienced cost when the flow is zero equals $b_j - s_j \geq L_\rho$, which means that $x_j^s = 0$. \square

Evidently, plugging the values into the expression of the previous theorem for Examples 1 and 2 gives us the rebates that we indicated in §3.1. What remains to be done to finish the characterization of optimal rebates is to find \mathcal{F}_0 , which will allow us to determine the value of D_ρ . The following result provides a characterization of the common cost experienced by participants under a Stackelberg equilibrium. We will use it to compute the values of D_ρ and V_ρ .

PROPOSITION 5.11. Consider a network with parallel arcs, affine cost functions, and total demand $r > 0$. For $j \in A$, define $\gamma(j, r) := (r + \sum_{i=1}^j (b_i/a_i))/K([j])$. There exist unique arcs $i_0, i_1 \in A$ such that

$$b_{i_0} \leq \gamma(i_0, r) < b_{i_0+1}, \quad (9a)$$

$$b_{i_1} < \gamma(i_1, r) \leq b_{i_1+1}. \quad (9b)$$

Moreover, $\gamma(i_0, r) = \gamma(i_1, r) = L_\infty$, where L_∞ is the common cost experienced by participants under a Wardrop equilibrium (without rebates).

PROOF. Let us define $i_0 := \max\{i \in A: b_i \leq L_\infty\}$, and let x be the Wardrop equilibrium. From the definition, i_0 satisfies that $b_{i_0} \leq L_\infty < b_{i_0+1}$. The equilibrium condition implies that $x_i = (L_\infty - b_i)/a_i$ for all $i \leq i_0$. Summing over that range we get that $L_\infty = \gamma(i_0, r)$. What is left to prove is that there is no other i_0 that satisfies (9a). Hence, assume that there is another index \tilde{i}_0 , and define \tilde{x} equal to $(\gamma(\tilde{i}_0, r) - b_i)/a_i$ for $i \leq \tilde{i}_0$ and 0 otherwise. This flow is feasible because it is nonnegative and its total demand equals r . Furthermore, it satisfies the Wardrop equilibrium conditions with cost equal to $\gamma(\tilde{i}_0, r)$ for all participants. Recall that because cost functions are strictly increasing, there exists a unique Wardrop equilibrium. Because x and \tilde{x} are both at equilibrium, they must be equal. This implies that $\gamma(i_0, r) = \gamma(\tilde{i}_0, r)$, from where $\tilde{i}_0 = i_0$ because of (9a).

A similar argument proves the existence of a unique index $i_1 := \max\{i \in A: b_i < L_\infty\}$ that satisfies (9b). We highlight that i_0 and i_1 differ only when there is a link i with $b_i = L_\infty$, in which case $i_0 > i_1$. \square

Computing $\gamma(i, r)$ for the different arcs in Instance 2, we get that $\gamma(1, r) = r$, $\gamma(2, r) = (r + 1)/2$, and $\gamma(3, r) = (r + 3)/3$. Then $i_0 = 1$ when $0 \leq r < 1$, $i_0 = 2$ when $1 \leq r < 3$, and $i_0 = 3$ when $r \geq 3$. Similarly, $i_1 = 1$ when $0 \leq r \leq 1$, $i_1 = 2$ when $1 < r \leq 3$, and $i_1 = 3$ when $r > 3$.

In the sequel, we consider the function $L_\infty(z)$ which represents the perceived cost under a Wardrop equilibrium (without rebates) when the total demand is z . When we do not denote a demand explicitly, we assume that the regular demand of r is used. It is well known that the function $L_\infty(z)$ is nondecreasing and continuous (Hall 1978). In addition, Proposition 5.11 implies that it is piecewise linear with slope $1/K([i])$ when its value is between b_i and b_{i+1} . Therefore, it is a concave function. For an illustration, see Figure 4 in the following section. Under our assumptions, $L_\infty(\cdot)$ is easy to compute using an incremental loading algorithm.

Using the previous result, we can now express the perceived cost of participants at the Stackelberg equilibrium. In addition, the next proposition will clearly identify the sets \mathcal{F}_0 and \mathcal{F}_s . First, $\mathcal{F}_0 = [i_0]$, where i_0 corresponds to the index introduced in Proposition 5.11 for a demand of $r(1 - 1/\rho)$. The arcs without rebates that are used in a Stackelberg equilibrium coincide with those that are used under a Wardrop equilibrium (without rebates) with a total demand of $r(1 - 1/\rho) + \epsilon$, for a sufficiently small $\epsilon > 0$. Likewise, $\mathcal{F}_s = [i_1]$, where i_1 is the index introduced in Proposition 5.11 for a demand of $r(1 + 1/\rho)$. The arcs used under a Stackelberg equilibrium coincide with those that are used under a Wardrop equilibrium with a total demand of $r(1 + 1/\rho)$.

PROPOSITION 5.12. Consider a network with parallel arcs and affine cost functions, and a Stackelberg equilibrium (s, x^s) . If rebates are beneficial (i.e., if $\mathcal{F}_s \neq \emptyset$), then

$$D_\rho = L_\infty\left(r \frac{\rho-1}{\rho}\right) \quad \text{and} \quad V_\rho = L_\infty\left(r \frac{\rho+1}{\rho}\right),$$

and the perceived cost of each participant under x^s is

$$L_\rho = \frac{1}{2} \left(L_\infty\left(r \frac{\rho-1}{\rho}\right) + L_\infty\left(r \frac{\rho+1}{\rho}\right) \right).$$

PROOF. From Proposition 5.8 and Lemma 5.9, we know that there exist $i_0, i_1 \in A$ such that $\mathcal{F}_0 = [i_0]$, $\mathcal{F} = [i_1]$,

$$b_{i_0} \leq \gamma\left(i_0, r \frac{\rho-1}{\rho}\right) < b_{i_0+1}, \quad \text{and}$$

$$b_{i_1} < \gamma\left(i_1, r \frac{\rho+1}{\rho}\right) \leq b_{i_1+1}.$$

Hence, Proposition 5.11 implies the first two claims. The third follows simply from the relation displayed right after (5). \square

Using the values of i_0 that we previously computed for Instance 2, it is easy to see that $L_\infty(r) = r$ when $0 \leq r < 1$, $L_\infty(r) = (r+1)/2$ when $1 \leq r < 3$, and $L_\infty(r) = (r+3)/3$ when $r \geq 3$. Using this, $D_\rho = 1 - 1/\rho$, $V_\rho = 1 + 1/(2\rho)$, and $L_\rho = 1 - 1/(4\rho)$ as expected.

Notice that Proposition 5.12 provides an explicit way to compute D_ρ . Hence, this value is unique and, relying on Proposition 5.11, the vector of optimal rebates is unique as well (disregarding that a rebate for an arc $l \in A \setminus \mathcal{F}$ can take any value between 0 and $c_l(0) - L_\rho$, which does not count as multiple equilibria because l is unused). Because there is a unique Wardrop equilibrium for any given rebate vector such that $L_\rho > 0$, the Stackelberg game has an essentially unique solution. This means that any two different Stackelberg equilibria will be undistinguishable from a practical point of view because flows and costs under both solutions will be equal.

The following proposition provides an easily verifiable condition to check whether rebates can help lower the social cost in a specific instance. Note that when the inequality does not hold, the formula must hold with equality because of the concavity of $L_\infty(r)$.

PROPOSITION 5.13. Consider a network with parallel arcs and affine cost functions. Rebates are beneficial (i.e., $\mathcal{F}_s \neq \emptyset$) if and only if

$$\frac{1}{2} \left(L_\infty\left(r \frac{\rho-1}{\rho}\right) + L_\infty\left(r \frac{\rho+1}{\rho}\right) \right) < L_\infty.$$

5.3. A Polynomial-Time Algorithm for Computing Optimal Rebates

The results we have presented in the previous section lead to a polynomial-time algorithm for finding the optimal rebates. The following algorithm receives an instance described by a network with parallel arcs, affine cost functions, and a fixed demand as input, and computes a Stackelberg equilibrium.

Algorithm

- (1) Sort the arcs with respect to b_i to cast the instance into the form we considered.
- (2) Compute the function $L_\infty(z)$ for the instance.
- (3) Use Proposition 5.13 to decide whether rebates need to be used.
- (4) If rebates are not beneficial, we are done.
- (5) Compute D_ρ using Proposition 5.12.
- (6) Finally, compute the rebate to offer in each arch using Theorem 5.10.

Each of these steps requires a computation that can be done in polynomial time. The bottleneck is computing $L_\infty(z)$, which requires solving at most $|A|$ systems of linear equations to load the network incrementally and compute the breakpoints of the piecewise linear function.

At this point, it is convenient to discuss how to estimate the information needed to create an instance in practice. This estimation has been discussed at length in the literature of transportation engineering (see, e.g., Sheffi 1985). We provide a short overview. First, one needs to list the modes and their costs as a function on the flows. Cost functions are calibrated from historical information, taking into account how different modes operate. Overall, one needs to sum the travel time and the fare or toll for the mode, which can be converted to the same units by using the average value of time for the population. The latter can usually be estimated from socioeconomic information coming from census data. The demand can be measured directly or may come from historical OD matrices that can be calibrated using up-to-date traffic counts. The most difficult parameter to estimate in our model is ρ because it is hard to attach a dollar figure to a reduction in the total cost experienced by travelers. This estimation has been attempted by the Partnership for New York City (2006), who measure the economic impact of reducing traffic congestion. Alternatively, one can compute the optimal rebates, social costs, and total cost experienced by commuters, as a function of ρ . This can be done easily because the algorithm above runs fast enough to solve the problem for many different values of ρ . With this curve in hand, one can look at the tradeoff between the budget invested in rebates and the overall social benefit. This can guide policymakers in selecting the optimal rebates to be used in a concrete situation.

6. The Benefits of Using Rebates

6.1. Coordination Mechanisms Based on Transfer Payments

First, we introduce some measures derived from the price of anarchy that will be useful to quantify the quality of equilibria resulting from a coordination mechanism. As we said in §2, Roughgarden and Tardos (2002) were the first to measure the price of anarchy in the network competition model introduced by Wardrop (1952). They defined the coordination ratio of an instance as

$$\frac{C(x^{\text{WE}})}{C(x^{\text{SO}})}, \quad (10)$$

and the price of anarchy as the supremum of (10) among all Wardrop equilibria and all possible instances (meaning all possible networks, demands, and allowed cost functions). Note that this value is at least 1 and it can be interpreted as follows: if it is low, then there is not much improvement to be expected from the introduction of a coordination mechanism in the game that was considered. On the other hand, a large price of anarchy suggests that there is a potentially large benefit to be made. For example, the coordination ratio of Pigou's instance (Instance 1) is 4/3. The following result establishes that this ratio is the largest possible.

PROPOSITION 6.1 (ROUGHGARDEN AND TARDOS 2002). *The price of anarchy for instances with affine cost functions is 4/3.*

For quadratic, cubic, and quartic cost functions, the price of anarchy is 1.626, 1.896, and 2.151, respectively (Roughgarden 2003; Correa, Schulz, and Stier-Moses 2004). For a simple proof of these results we direct the reader to Correa, Schulz, and Stier-Moses (2008).

Traditionally, the efficiency of a solution involving congestion pricing has been defined in terms of the total cost $C(\cdot)$ because charges are transfer payments that stay inside the system, or alternatively by assuming that these payments can be redistributed back to the users. For that social cost function, as Proposition 3.5 shows, charging users the externalities they introduce produces a socially efficient outcome. Some more recent articles look at social cost functions that include a term corresponding to taxation, similar to what we do in (1a). Under these more general social cost functions, a system owner may take a more holistic view, and not only care about outcomes, but also about investments. Cole, Dodis, and Roughgarden (2006) considered the problem of finding the taxes τ that minimize $\sum_{i \in A} x_i(c_i(x_i) + \tau_i)$, where x is a Wardrop equilibrium with respect to modified cost functions $c(\cdot) + \tau$. Unfortunately, finding an optimal mechanism for this

social cost function is NP-hard for arbitrary instances. Although they did not explicitly specialize their results to networks with parallel arcs, a generalization of the results of §5 can be used to compute optimal payments in polynomial time (still considering a general conversion factor ρ like in (1a)). Karakostas and Kolliopoulos (2005) extended the previous analysis and found bounds for the social cost achieved by an extension of the marginal taxation mechanism to heterogeneous values-of-time. Under this setting, the ratio of the social cost of an equilibrium to the solution of minimum social cost with respect to the optimal taxes is bounded with a smaller constant than that of Proposition 6.1 and its generalizations. In addition, the social cost is not too large compared to the minimum possible total cost (without taxes).

One can use different variations of the concept of the price of anarchy to quantify the power of a coordination mechanism. We consider the two definitions that are most interesting in our opinion. Both consist of ratios of the same cost function under two different solutions, thereby not falling into the situation of comparing apples and oranges. In addition, both compare the outcome provided by the coordination mechanism to an upper or lower bound, depending on the circumstances.

The first measure we consider is a straightforward extension of (10). Indeed, to quantify the loss of efficiency because of the limited coordinating power of the system owner, we consider the ratio

$$\frac{C(x^s)}{C(x^{\text{SO}})}. \quad (11)$$

For example, looking at Instance 2, this ratio equals 1 for $\rho \leq 1$ and $(8 - 2/\rho + 1/\rho^2)/7$ for $\rho > 1$. Note that although the previous ratio measures the quality of a given solution (s, x^s) for a fixed instance, our main interest is on the supremum of the coordination ratio of an arbitrary Stackelberg equilibrium over all possible instances, as it is done for the price of anarchy. Another option would have been to define the price of anarchy as in (11) but using perceived costs, as Cole, Dodis, and Roughgarden (2006) proposed for their study of taxes in networks. Remark 6.5 shows that the bound that can be obtained is the same as that for (11).

Previous research has determined that the price of anarchy is sometimes a pessimistic measure, as can be expected from a general worst-case bound. For example, Correa, Schulz, and Stier-Moses (2008) proposed to restrict the analysis to instances with fixed congestion loads to get more realistic estimates. Another aspect of the previous definitions is that they do not consider that in certain settings a system optimum is unrealistic and cannot be implemented. For example, Schulz and Stier-Moses (2006) proposed

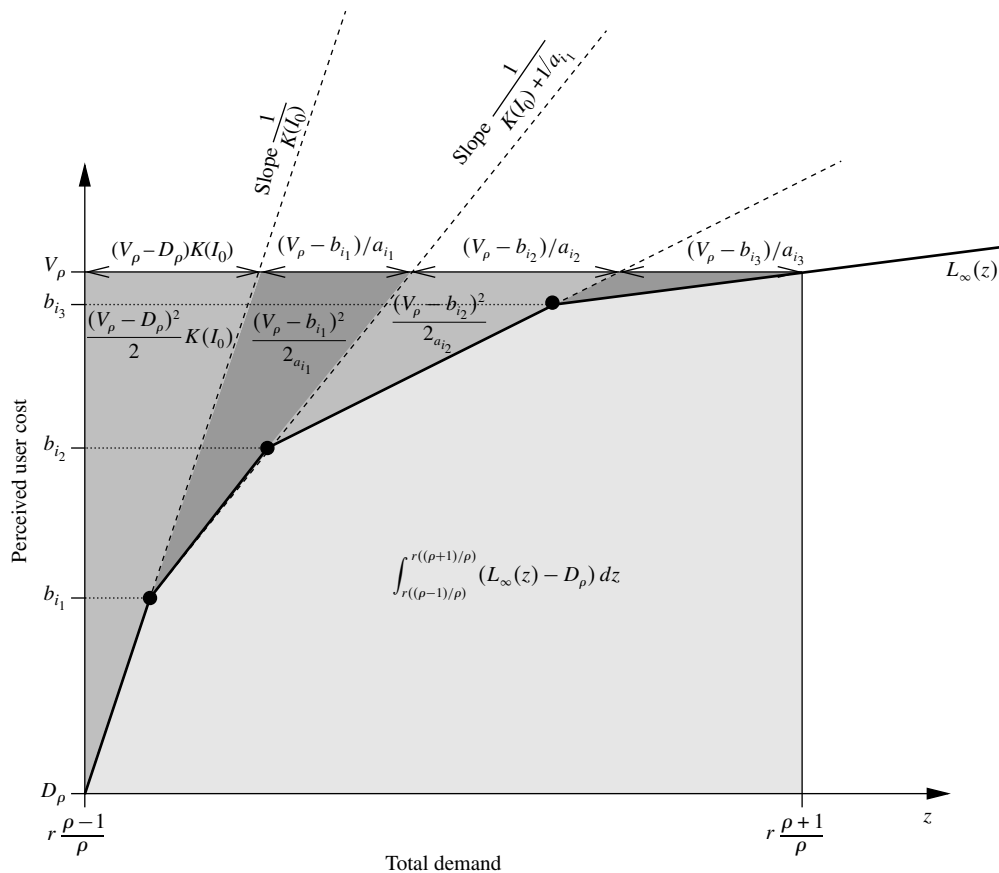


Figure 4 Illustration of the Decomposition of $2r(V_\rho - D_\rho)/\rho$

to quantify the performance of a route guidance system for vehicular traffic by comparing the solutions with and without guidance instead of using a social optimum.

To get a measure that is both less pessimistic and more realistic, we consider that the best possible outcome is what the system owner can enforce by setting rebates correctly. Hence, we consider the ratio of the social cost of a Wardrop equilibrium to that of a Stackelberg equilibrium. Letting s be the optimal rebate vector, this ratio is expressed as

$$\frac{C_\rho(0)}{C_\rho(s)} = \frac{C(x^{WE})}{C_\rho(s)}. \tag{12}$$

When $\rho < 1$, this quantity may be large because the denominator of (12) can be arbitrary small. Instead, when $\rho \geq 1$, the lower bound in (2) implies that this ratio is less pessimistic (smaller) than (10). For the examples provided before, we get that this ratio equals $4\rho/(4\rho - 1)$ for Instance 1, while the coordination ratio displayed in (10) equals $4/3$. The corresponding values for Instance 2 are $8\rho/(8\rho - 1)$ and $8/7$, respectively.

6.2. Computing the Price of Anarchy

Now that we have already characterized the optimal rebates for a particular instance of the problem, we are ready to analyze the performance of this coordination mechanism. We continue to work with networks consisting of parallel arcs and affine cost functions.

We start by providing a bound between the uncoordinated solution (no rebates) and the Stackelberg equilibrium. The case of $\rho \leq 1$ follows from Proposition 4.2. Indeed, using Proposition 6.1, we have that $C_\rho(0)/C_\rho(s) = C(x^{WE})/(\rho C(x^{SO})) \leq 4/(3\rho)$. This means that the price of anarchy arising from (12) is $4/(3\rho)$ for an arbitrary network with affine cost functions. The case of $\rho > 1$ is more involved. We start by computing the social cost of the Stackelberg equilibrium making use of the relations developed in the previous section.

LEMMA 6.2. Consider a network with parallel arcs and affine cost functions. For $\rho > 1$, the optimal social cost equals $(\rho/2) \int_{r(1-1/\rho)}^{r(1+1/\rho)} L_\infty(z) dz$.

PROOF. We rewrite the expression $2r(V_\rho - D_\rho)/\rho$ using the graphical decomposition shown in Figure 4.

Indeed, the area of the rectangle equals

$$\begin{aligned} & K(\mathcal{J}_0) \frac{(V_\rho - D_\rho)^2}{2} + \sum_{i \in \mathcal{J}_s} \frac{(V_\rho - b_i)^2}{2a_i} + \int_{r(1-1/\rho)}^{r(1+1/\rho)} (L_\infty(z) - D_\rho) dz \\ &= \frac{V_\rho - D_\rho}{2} \underbrace{\left((V_\rho - D_\rho)K(\mathcal{J}_0) + \sum_{i \in \mathcal{J}_s} \frac{V_\rho - b_i}{a_i} \right)}_{=2r/\rho} \\ & \quad - 2 \sum_{i \in \mathcal{J}_s} x_i^s s_i + \int_{r(1-1/\rho)}^{r(1+1/\rho)} (L_\infty(z) - D_\rho) dz, \end{aligned}$$

where we used the expression for x_i^s in the proof of Lemma 5.9, the expression for s_i in Theorem 5.10, and that $(V_\rho - b_i)^2 = (V_\rho - b_i)(V_\rho - D_\rho + D_\rho - b_i)$. The term with the brace equals $2r/\rho$ because of (17). After some algebra,

$$\begin{aligned} \sum_{i \in \mathcal{J}_s} x_i^s s_i &= \frac{1}{2} \int_{r(1-1/\rho)}^{r(1+1/\rho)} (L_\infty(z) - D_\rho) dz - \frac{r}{2\rho} (V_\rho - D_\rho) \\ &= \frac{1}{2} \int_{r(1-1/\rho)}^{r(1+1/\rho)} L_\infty(z) dz - \frac{r}{\rho} L_\rho. \end{aligned} \tag{13}$$

Consequently, the optimal social cost is

$$\begin{aligned} C_\rho(s) &= rL_\rho + \rho \sum_{i \in \mathcal{J}_s} x_i^s s_i \\ &= (\rho/2) \int_{r(1-1/\rho)}^{r(1+1/\rho)} L_\infty(z) dz. \quad \square \end{aligned} \tag{14}$$

THEOREM 6.3. Consider a network with parallel arcs and affine cost functions. For $\rho > 1$, the unique Stackelberg equilibrium (s, x^s) satisfies that

$$\frac{C_\rho(0)}{C_\rho(s)} \leq \frac{4\rho}{4\rho - 1}.$$

PROOF. Let us assume that $\mathcal{J}_s \neq \emptyset$ because otherwise the result is trivial. We need to compare the cost $C_\rho(s)$ computed in the previous lemma to $rL_\infty(r)$. Since $L_\infty(z)$ is a positive and concave function, $L_\infty(z)/z$ is a nonincreasing function. Bounding the integral from below as Figure 5 illustrates, we get that

$$C_\rho(s) \geq \frac{\rho r L_\infty(r)}{2\rho} \left(2 - \frac{1}{2\rho} \right) = C_\rho(0) \left(1 - \frac{1}{4\rho} \right),$$

as claimed. \square

The previous result characterizes the tradeoff between willingness to offer rebates and coordination power of the mechanism. The corresponding bound is tight, as Instance 1 demonstrates. (Note that the topmost arc has a constant cost, but one can take that cost equal to $ax + 1$ for an arbitrarily small a and nothing changes.) When the system owner’s willingness to offer rebates is high (ρ is not much larger than 1),

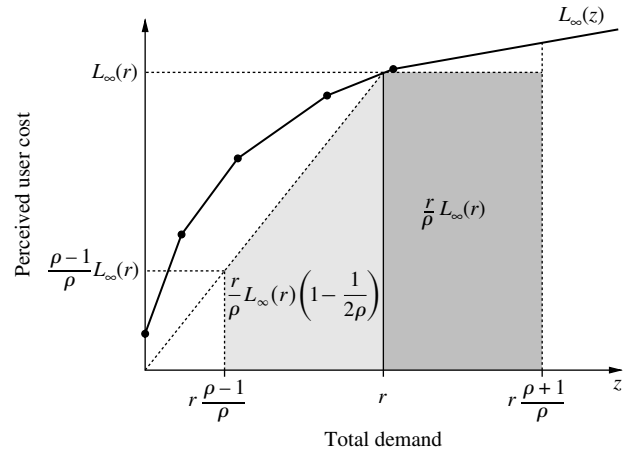


Figure 5 Illustration of the Bound for $\int_{r(1-1/\rho)}^{r(1+1/\rho)} L_\infty(z) dz$

the optimal social cost is approximately equal to the total cost under a system optimum; hence, the previous theorem provides a bound that is close to $4/3$. Here, recall that $4/3$ is the price of anarchy when the coordination mechanism can achieve a socially optimal solution (Proposition 6.1). Not surprisingly, when the willingness to offer rebates decreases (big ρ), the previous theorem gives a bound that is close to 1 because the system owner cannot do much better than in a Wardrop equilibrium.

Finally, we compute the worst-case ratio between the participants’ real cost under a Stackelberg equilibrium and under a system optimum, as we proposed in (11). In the case of $\rho \leq 1$, the flow x^{SO} is at equilibrium (and it is the unique one for strictly increasing cost functions, see §4.1), which implies that for an arbitrary network with affine cost functions the mechanism coordinates the network. The following results provide the bound corresponding to the case of $\rho > 1$.

THEOREM 6.4. Consider a network with parallel arcs and affine cost functions. The Stackelberg equilibrium (s, x^s) described in the previous section satisfies that

$$\frac{C(x^s)}{C(x^{SO})} \leq \frac{4\rho}{3\rho + 1}. \tag{15}$$

This bound is close to 1 for $\rho \approx 1$ because in that case a Stackelberg equilibrium is similar to a system optimum, and close to $4/3$ when ρ is large because in that case it is similar to a Wardrop equilibrium. As for the previous bound, Theorem 6.4 provides the curve that characterizes the tradeoff between willingness to offer rebates and coordinating power. We highlight that this bound is tight, which can be observed by taking $\epsilon = 0$ and letting α tend to 0 in Instance 3.

REMARK 6.5. The bound provided by Theorem 6.4 is also valid if one takes the ratio of the participants’ perceived cost in the Stackelberg equilibrium to that in the system optimum. This holds

because $\sum_i x_i^s [c_i(x_i^s) - s_i]^+ \leq C(x^s)$. Moreover, the same instance as before shows that this bound is tight.

7. Conclusions

We have studied the possible improvement that can stem from the use of rebates to coordinate an urban transportation network. If a system owner can afford to offer rebates and the system is highly congestible, rebates can significantly lower the social cost, which includes commute times and costs, as well as the cost of providing the rebates themselves. The algorithm we have presented can be used to determine optimal subsidies for each mode of transportation. Subsidies only affect a limited proportion of the demand, implying that the cost of providing them will not be exceedingly large. We have also estimated how much improvement this coordination mechanism brings to the system, as a function of the city’s sensitivity to the cost of offering rebates. The coordinating power of a rebate scheme increases as the owner’s sensitivity to the rebate cost decreases.

Several questions related to this study remain open. First and foremost, we have worked under the assumption that an instance has parallel arcs and affine cost functions. It would be interesting to generalize our results to more general instances. Another interesting problem is to determine the computational complexity of finding optimal rebates. Proving its hardness would shed light into this problem and would motivate the need to look for good heuristics. For quadratic cost functions for example, optimal rebates can be irrational numbers.⁴ Hence, an optimal rebate vector cannot be computed exactly in polynomial time. Nevertheless, it would be interesting to find a way to approximate it. Finally, another interesting open question is whether optimal rebates are unique in general. We have shown that this is true for networks with parallel arcs and affine cost functions.

Our model has some limitations that we would like to address in future research. On the one hand, we plan to incorporate the possibility that the system owner considers congestion pricing and rebates at the same time. Such extension will be useful to model systems in which both incentive mechanisms co-exist to create a larger differential between the total cost of driving and that of public transportation. On the other hand, we also want to look at an heterogeneous population because the valuation of time is user-dependent. This extension would allow a modeler to look at more precise measures of equity among commuters. Furthermore, it is important to consider

elastic demands because in practice some trips are optional and will not happen if the price of transportation is too high. The last element that would be interesting to consider is a situation in which multiple agencies in the government have to coordinate their efforts and budgets to offer incentives to the population. Because each agency has its own goal and agenda, they may not agree in the policy that should be chosen.

Acknowledgments

The authors thank two anonymous referees for their constructive remarks that improved the presentation of this paper, Tom Liebling for stimulating discussions about the motivations of this paper, and participants of the Mittagsemnar in the Operations Research Center at MIT for their insightful comments. This work has been supported by the Center for International Business Education and Research (CIBER), and by the Center for Excellence in E-Business (CEBIZ), both at Columbia University. The research was done while the first author was visiting Columbia Business School.

Appendix A. Proofs

A.1. Proof of Lemma 5.2

PROOF. Let $\tilde{A} := A \setminus \{l\}$ and \tilde{s} be the restriction of s to \tilde{A} . Assume that \tilde{s} is not optimal for \tilde{A} , and let \tilde{s}^* be an optimal rebate vector for that network. Then, $C_{\rho}^{\tilde{A}}(\tilde{s}^*) < C_{\rho}^{\tilde{A}}(\tilde{s}) = C_{\rho}^A(s)$, where the superscript represents the instance and the equality holds because if no participant selects the arc, it makes no difference whether the arc exists. Now, we take the optimal rebate vector \tilde{s}^* and extend it to the original network by setting $s_l^* := 0$ and $s_i^* := \tilde{s}_i^*$ for $i \in \tilde{A}$. Because a situation like Braess’ paradox (1968) cannot occur in networks with parallel arcs, the participants’ real cost at a Wardrop equilibrium decreases when link l is (re)introduced. Together with the fact that arc l is not subsidized, we have that $C_{\rho}^A(s^*) \leq C_{\rho}^A(\tilde{s}^*)$, which contradicts the optimality of s in the original instance. \square

A.2. Proof of Proposition 5.3

PROOF. Without loss of generality assume that $s_i = c_i(0) - L_{\rho}$ for all $i \in A \setminus \mathcal{F}$. Consider two fixed arcs $i \in \mathcal{F}_s$ and $j \in A$. Since s_i is strictly positive, it is possible to simultaneously reduce s_i by a positive infinitesimal ds_i and increase s_j so that the only effect is that some participants switch from arc i to j . In other words, we have that $dx_j = -dx_i$, where we denote an infinitesimal variation of a quantity w by dw . By design, the perceived cost L_{ρ} at equilibrium remains the same. The local effect at the arcs in question is $d(c_i(x_i^s) - s_i) = 0$ and $d(c_j(x_j^s) - s_j) = 0$. Because s was optimal, this modification cannot decrease the total rebate cost $\sum_{i \in A} x_i^s s_i$, as it does not modify the total participants’ perceived cost. This implies that $d(x_i^s s_i + x_j^s s_j) \geq 0$. Putting all together,

$$dx_i^s (x_i^s c_i'(x_i^s) + s_i - x_j^s c_j'(x_j^s) - s_j) \geq 0.$$

As $dx_i^s < 0$, we must have $x_i^s c_i'(x_i^s) + s_i \leq x_j^s c_j'(x_j^s) + s_j$, and adding $c_i(x_i^s) - s_i = c_j(x_j^s) - s_j$, we finally obtain that $c_i(x_i^s) + x_i^s c_i'(x_i^s) \leq c_j(x_j^s) + x_j^s c_j'(x_j^s)$. We get the claim by letting i and j vary. \square

⁴ For example, considering the instance shown in Figure 1 with costs functions 1 and x^2 , and $\rho = 2$, it is optimal to offer a rebate of $(11 - \sqrt{13})/18$ for the arc with constant cost.

A.3. Proof of Lemma 5.5

PROOF. To ensure that the modification to the rebates we are going to make does not change the sets \mathcal{F}_s and \mathcal{F}_0 , we first remove all unused arcs. Indeed, Lemma 5.2 proves that if s is optimal for the original network, it is also optimal for the instance containing the arcs in \mathcal{F} only. The proposition is obvious for $\mathcal{F}_s = \emptyset$, so let us assume the opposite. We consider adding or subtracting a common infinitesimal ds to all rebates that are strictly positive. After modifying s the outcome is still at equilibrium and all arcs are still used; hence, differentials of perceived costs are equal for all arcs in \mathcal{F} . For a fixed $i_s \in \mathcal{F}_s$ and a fixed $i_0 \in \mathcal{F}_0 \neq \emptyset$, we have that $c'_i(x_i^s)dx_{i_s} - ds = c'_{i_0}(x_{i_0}^s)dx_{i_0}^s$ and

$$dx_i^s = \begin{cases} c'_i(x_i^s)dx_{i_s}^s / c'_i(x_i^s) & i \in \mathcal{F}_s, \\ c'_{i_0}(x_{i_0}^s)dx_{i_0}^s / c'_i(x_i^s) & i \in \mathcal{F}_0. \end{cases}$$

As the total demand does not change, we must have that $0 = \sum_{i \in \mathcal{F}} dx_i^s = K(\mathcal{F}_s)c'_i(x_i^s)dx_{i_s}^s + K(\mathcal{F}_0)c'_{i_0}(x_{i_0}^s)dx_{i_0}^s$. After some algebra, $c'_i(x_i^s)dx_{i_s}^s = dsK(\mathcal{F}_0)/K(\mathcal{F})$. Finally, let us consider how the social cost changes.

$$\begin{aligned} dC_\rho(s) &= d\left(r(c_{i_s}(x_{i_s}) - s_{i_s}) + \rho \sum_{i \in \mathcal{F}} x_i s_i\right) \\ &= r(c'_{i_s}(x_{i_s}^s)dx_{i_s}^s - ds) + \rho \sum_{i \in \mathcal{F}_s} \left(c'_{i_s}(x_{i_s}^s)dx_{i_s}^s \frac{s_i}{c'_i(x_i^s)} + x_i^s ds\right) \\ &= c'_{i_s}(x_{i_s}^s)dx_{i_s}^s \left(r + \rho \sum_{i \in \mathcal{F}_s} \frac{s_i}{c'_i(x_i^s)}\right) + ds\left(\rho \sum_{i \in \mathcal{F}_s} x_i^s - r\right) \\ &= ds\left(\left(\frac{K(\mathcal{F}_0)}{K(\mathcal{F})} - 1\right)r + \rho \frac{K(\mathcal{F}_0)}{K(\mathcal{F})} \sum_{i \in \mathcal{F}_s} \frac{s_i}{c'_i(x_i^s)} + \rho \sum_{i \in \mathcal{F}_s} x_i^s\right) \\ &= ds \frac{\rho}{K(\mathcal{F})} \left(-\frac{r}{\rho}K(\mathcal{F}_s) + K(\mathcal{F}_0) \sum_{i \in \mathcal{F}_s} \frac{s_i}{c'_i(x_i^s)} + K(\mathcal{F}) \sum_{i \in \mathcal{F}_s} x_i^s\right). \end{aligned}$$

The claim follows because the optimality of s implies that $dC_\rho(s) \geq 0$ for feasible directions $ds > 0$ and $ds < 0$. \square

A.4. Proof of Lemma 5.9

PROOF. On the one hand, (3a) and (4a), respectively, imply that

$$x_i^s = \begin{cases} \frac{V_\rho + D_\rho - 2b_i}{2a_i} & i \in \mathcal{F}_0, \\ \frac{V_\rho - b_i}{2a_i} & i \in \mathcal{F}_s. \end{cases}$$

Since $\sum_{i \in \mathcal{F}} x_i^s = r$, we have

$$\frac{V_\rho}{2}K(\mathcal{F}) = r - \frac{D_\rho}{2}K(\mathcal{F}_0) + \sum_{i \in \mathcal{F}_s} \frac{b_i}{2a_i} + \sum_{i \in \mathcal{F}_0} \frac{b_i}{a_i}. \quad (16)$$

On the other hand, (4a), (5a), and Lemma 5.5 imply that

$$\begin{aligned} \frac{r}{\rho}K(\mathcal{F}_s) &= \frac{V_\rho}{2}K(\mathcal{F}_s)K(\mathcal{F}) + (K(\mathcal{F}_0) - K(\mathcal{F})) \sum_{i \in \mathcal{F}_s} \frac{b_i}{2a_i} \\ &\quad - \frac{D_\rho}{2}K(\mathcal{F}_0)K(\mathcal{F}_s), \end{aligned}$$

and since $\mathcal{F}_s \neq \emptyset$,

$$\frac{V_\rho}{2}K(\mathcal{F}) = \frac{r}{\rho} + \sum_{i \in \mathcal{F}_s} \frac{b_i}{2a_i} + \frac{D_\rho}{2}K(\mathcal{F}_0). \quad (17)$$

Adding and subtracting (16) and (17) yield the claim. \square

A.5. Proof of Proposition 5.13

PROOF. Rebates are beneficial only if the social cost of a Stackelberg equilibrium is lower than that of the Wardrop equilibrium. In that case, $L_\rho < L_\infty$ and, hence, the strict inequality of the claim holds because of Proposition 5.12.

We now focus on the reverse implication. Assuming that the inequality in the hypothesis holds, there exists an $i \in A$ such that

$$L_\infty \left(r \frac{\rho - 1}{\rho}\right) < b_i < L_\infty \left(r \frac{\rho + 1}{\rho}\right).$$

Proposition 5.8 implies that if $i \in A \setminus \mathcal{F}$ then $b_i \geq V_\rho$, and if $i \in \mathcal{F}_0$ then $b_i \leq D_\rho$. Therefore, $i \in \mathcal{F}_s$, which is consequently nonempty. \square

A.6. Proof of Theorem 6.4

First, we express the cost of the system optimum as a function of L_∞ to be able to relate it to a Stackelberg equilibrium.

LEMMA A.1. For networks with parallel arcs and affine cost functions, the minimal value of the participants' real cost is

$$C(x^{\text{SO}}) = \frac{1}{2} \int_{z=0}^{2r} L_\infty(z) dz.$$

PROOF. Proposition 3.5 implies that there exists a constant $L^{\text{SO}} > 0$ such that

$$\begin{aligned} L^{\text{SO}} &= 2a_i x_i^{\text{SO}} + b_i \quad \forall i \text{ s.t. } x_i^{\text{SO}} > 0, \\ L^{\text{SO}} &\leq b_i \quad \forall i \text{ s.t. } x_i^{\text{SO}} = 0. \end{aligned}$$

Proceeding as in Stackelberg equilibrium case, $x_i^{\text{SO}} = [L^{\text{SO}} - b_i]^+ / (2a_i)$. If we let $i^{\text{SO}} := \max\{i \in A : b_i \leq L^{\text{SO}}\}$, we have that $r = \sum_{j=1}^{i^{\text{SO}}} (L^{\text{SO}} - b_j) / (2a_j)$. Hence,

$$L^{\text{SO}} = \frac{1}{K([i^{\text{SO}}])} \left(2r + \sum_{j=1}^{i^{\text{SO}}} \frac{b_j}{a_j}\right).$$

Since $b_{i^{\text{SO}}} \leq L^{\text{SO}} < b_{i^{\text{SO}}+1}$, Proposition 5.11 implies that $L^{\text{SO}} = L_\infty(2r)$. Then

$$\begin{aligned} C(x^{\text{SO}}) &= \sum_{j=1}^{i^{\text{SO}}} x_j^{\text{SO}} \left(L_\infty(2r) - \frac{L_\infty(2r) - b_j}{2}\right) \\ &= rL_\infty(2r) - \frac{1}{2} \sum_{j=1}^{i^{\text{SO}}} \frac{(L_\infty(2r) - b_j)^2}{2a_j}. \end{aligned}$$

Finally, using a similar decomposition as in Figure 4, it can be shown that

$$2rL_\infty(2r) = \int_{z=0}^{2r} L_\infty(z) dz + \sum_{i \in A : b_i \leq L_\infty(2r)} \frac{(L_\infty(2r) - b_i)^2}{2a_i}.$$

The claim follows from the last two equations. \square

Now we are ready to offer the proof of Theorem 6.4.

PROOF. Equations (1b), (13), and (14) imply that

$$\sum_{i \in A} x_i^s c_i(x_i^s) = r \frac{\rho - 1}{\rho} L_\rho + \frac{1}{2} \int_{r(1-1/\rho)}^{r(1+1/\rho)} L_\infty(z) dz. \quad (18)$$

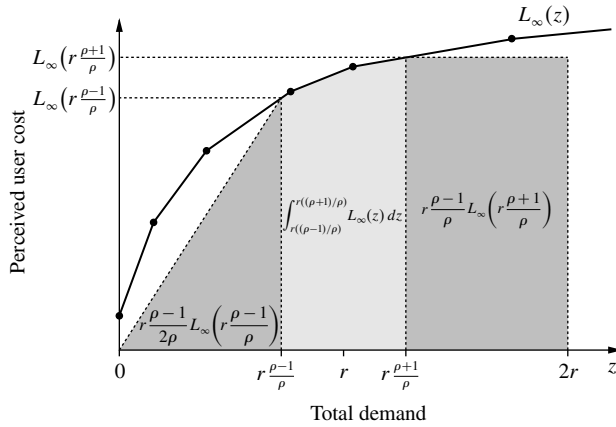


Figure A.1 Illustration of the Bound for $C(x^{SO})$

From Lemma A.1, the concavity of L_{∞} , and decomposing the area under the curve as Figure A.1 illustrates, we have that

$$\begin{aligned} 2C(x^{SO}) &\geq r \frac{\rho-1}{\rho} \frac{D_{\rho}}{2} + r \frac{\rho-1}{\rho} V_{\rho} + \int_{r(1-1/\rho)}^{r(1+1/\rho)} L_{\infty}(z) dz \\ &= 2 \sum_{i \in A} x_i^s c_i(x_i^s) - r \frac{\rho-1}{\rho} \frac{D_{\rho}}{2} \\ &\geq 2 \sum_{i \in A} x_i^s c_i(x_i^s) - r \frac{\rho-1}{\rho} \frac{L_{\rho}}{2} \\ &\geq C(x^s) \left(2 - \frac{\rho-1}{2\rho} \right), \end{aligned}$$

where the second, third, and fourth lines hold because of (18), $D_{\rho} \leq L_{\rho}$, and $rL_{\rho} \leq \sum_i x_i^s c_i(x_i^s)$, respectively. \square

References

Acemoglu, D., A. Ozdaglar. 2007. Competition and efficiency in congested markets. *Math. Oper. Res.* **32**(1) 1–31.

Ali, A., M. A. Jolson, R. Y. Darmon. 1994. A model for optimizing the refund value in rebate promotions. *J. Bus. Res.* **29**(3) 239–245.

Beckmann, M. J., C. B. McGuire, C. B. Winsten. 1956. *Studies in the Economics of Transportation*. Yale University Press, New Haven, CT.

Bergendorff, P., D. W. Hearn, M. V. Ramana. 1997. Congestion toll pricing of traffic networks. P. M. Pardalos, D. W. Hearn, W. W. Hager, eds. *Network Optimization. Lecture Notes in Economics and Mathematical Systems*, Vol. 450. Springer, Berlin, 51–71.

Braess, D. 1968. Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung* **12** 258–268. [Translated by: D. Braess, A. Nagurney, T. Wakolbinger. 2005. On a paradox of traffic planning. *Transportation Sci.* **39**(4) 446–450.]

Button, K. J., E. T. Verhoef. 1998. *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility*. Edward Elgar, Cheltenham, UK.

Chau, C. K., K. M. Sim. 2003. The price of anarchy for nonatomic congestion games with symmetric cost maps and elastic demands. *Oper. Res. Lett.* **31**(5) 327–334.

Chen, X., C.-L. Li, D. Simchi-Levi. 2007. The impact of manufacturer rebates on supply chain profits. *Naval Res. Logist.* **54**(6) 667–680.

Cole, R., Y. Dodis, T. Roughgarden. 2006. How much can taxes help selfish routing? *J. Comput. System Sci.* **72**(3) 444–467.

Correa, J. R., A. S. Schulz, N. E. Stier-Moses. 2004. Selfish routing in capacitated networks. *Math. Oper. Res.* **29**(4) 961–976.

Correa, J. R., A. S. Schulz, N. E. Stier-Moses. 2008. A geometric approach to the price of anarchy in nonatomic congestion games. *Games Econ. Behav.* **64** 457–469.

de Palma, A., N. Picard. 2006. Equilibria and information provision in risky networks with risk-averse drivers. *Transportation Sci.* **40**(4) 393–408.

de Palma, A., M. Kilani, R. Lindsey. 2007. Maintenance, service quality and congestion pricing with competing roads. *Transportation Res.* **41B**(5) 573–591.

Dupuit, J. 1849. De l’Influence des Péages sur l’Utilities des Voies de Communication. *Annales des Ponts et Chaussées* **17**. [Translated by: E. Henderson, 1962, as: *On Tolls and Transport Charges. International Economic Papers*, Vol. 11. Macmillan, London, 7–31.]

Engel, E., R. Fischer, A. Galetovic. 2004. Toll competition among congested roads. *Topics Econ. Anal. Policy* **4**(1).

Friedman, E. J. 2004. Genericity and congestion control in selfish routing. *Proc. 43rd Conf. Decision and Control*, Institute of Electrical and Electronics Engineers, Paradise Island, Bahamas, 4667–4672.

Gerstner, E., J. D. Hess. 1991. Who benefits from large rebates: Manufacturer, retailer or consumer? *Econom. Lett.* **36** 5–8.

Golany, B., U. G. Rothblum. 2006. Inducing coordination in supply chains through linear reward schemes. *Naval Res. Logist.* **53**(1) 1–15.

Hall, M. A. 1978. Properties of the equilibrium state in transportation networks. *Transportation Sci.* **12**(3) 208–216.

Haurie, A., P. Marcotte. 1985. On the relationship between Nash-Cournot and Wardrop equilibria. *Networks* **15**(3) 295–308.

Johari, R., J. N. Tsitsiklis. 2004. Network resource allocation and a congestion game. *Math. Oper. Res.* **29**(3) 407–435.

Johari, R., G. Y. Weintraub, B. Van Roy. 2009. Investment and market structure in industries with congestion. *Oper. Res.* Forthcoming.

Karakostas, G., S. G. Kolliopoulos. 2005. The efficiency of optimal taxes. A. López-Ortiz, A. Hamel, eds. *Combinatorial and Algorithmic Aspects of Networking (CAAN’04)*, Banff, Canada. *Lecture Notes in Computer Science*, Vol. 3405. Springer, Heidelberg, 3–12.

Kheel, T. W. 2008. Balancing free transit and congestion pricing in New York City. Report, Nurture New York’s Nature, New York.

Knight, F. H. 1924. Some fallacies in the interpretation of social cost. *Quart. J. Econom.* **38** 582–606.

Korilis, Y., A. Lazar, A. Orda. 1995. Architecting noncompetitive networks. *IEEE J. Selected Areas Comm.* **13**(7) 1241–1251.

Koutsoupias, E. 2004. Coordination mechanisms for congestion games. *ACM SIGACT News* **35**(4) 58–71.

Koutsoupias, E., C. H. Papadimitriou. 1999. Worst-case equilibria. C. Meinel, S. Tison, eds. *Proc. 16th Annual Sympos. Theoret. Aspects Comput. Sci. (STACS)*, Trier, Germany. *Lecture Notes in Computer Science*, Vol. 1563. Springer, Heidelberg, 404–413.

Labbé, M., P. Marcotte, G. Savard. 1998. A bilevel model of taxation and its application to optimal highway pricing. *Management Sci.* **44**(12) 1608–1622.

Luo, Z.-Q., J.-S. Pang, D. Ralph. 1996. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, Cambridge, UK.

Mas-Colell, A., M. D. Whinston, J. R. Green. 1995. *Microeconomic Theory*. Oxford University Press, New York.

Papadimitriou, C. H. 2001. Algorithms, games, and the Internet. *Proc. 33rd Annual ACM Sympos. Theory Comput. (STOC)*, Hersonissos, Greece, ACM Press, New York, 749–753.

Partnership for New York City. 2006. Growth or gridlock? The economic case for traffic relief and transit improvement for a greater New York, New York. Report, Partnership for New York City, New York.

Perakis, G. 2007. The “price of anarchy” under nonlinear and asymmetric costs. *Math. Oper. Res.* **32**(3) 614–628.

Perakis, G., G. Roels. 2007. The price of anarchy in supply chains: Quantifying the efficiency of price-only contracts. *Management Sci.* **53**(8) 1249–1268.

Pigou, A. C. 1920. *The Economics of Welfare*. Macmillan, London.

Qiu, L., Y. R. Yang, Y. Zhang, S. Shenker. 2006. On selfish routing in Internet-like environments. *Trans. Networking* **14**(4) 725–738.

- Rosenthal, R. W. 1973. A class of games possessing pure-strategy Nash equilibria. *Internat. J. Game Theory* **2**(1) 65–67.
- Roughgarden, T. 2003. The price of anarchy is independent of the network topology. *J. Comput. System Sci.* **67**(2) 341–364.
- Roughgarden, T. 2004. Stackelberg scheduling strategies. *SIAM J. Comput.* **33**(2) 332–350.
- Roughgarden, T. 2006. On the severity of Braess's paradox: Designing networks for selfish users is hard. *J. Comput. System Sci.* **72**(5) 922–953.
- Roughgarden, T., É. Tardos. 2002. How bad is selfish routing? *J. ACM* **49**(2) 236–259.
- Santos, G. 2005. Urban congestion charging: A comparison between London and Singapore. *Transport Rev.* **25**(5) 511–534.
- Santos, G., G. Fraser. 2006. Road pricing: Lessons from London. *Econom. Policy* **21**(46) 263–310.
- Schrank, D., T. Lomax. 2007. The 2007 urban mobility study. Texas Transportation Institute. Accessed September 13, 2009, <http://mobility.tamu.edu/ums>.
- Schulz, A. S., N. E. Stier-Moses. 2006. Efficiency and fairness of system-optimal routing with user constraints. *Networks* **48**(4) 223–234.
- Sheffi, Y. 1985. *Urban Transportation Networks*. Prentice-Hall, Englewood, NJ.
- Smith, M. J. 1979. The existence, uniqueness and stability of traffic equilibria. *Transportation Res.* **13B**(4) 295–304.
- Starkie, D. 1986. Efficient and politic congestion tolls. *Transportation Res.* **20A**(2) 169–173.
- Taylor, T. 2002. Supply chain coordination under channel rebates with sales effort effects. *Management Sci.* **48**(8) 992–1007.
- Vickrey, W. S. 1955. A proposal for revising New York's subway fare structure. *J. Oper. Res. Soc. America* **3**(1) 38–68.
- Vickrey, W. S. 1969. Congestion theory and transport investment. *Amer. Econom. Rev.* **59** 251–261.
- von Stackelberg, H. 1934. *Marktform und Gleichgewicht*. Springer, Wien. [Translated by: A. T. Peacock. 1952. *The Theory of the Market Economy*. Oxford University Press, New York].
- Wardrop, J. G. 1952. Some theoretical aspects of road traffic research. *Proc. Institution of Civil Engineers, Part II*, Vol. 1, 325–378.
- Wichiensin, M., M. G. H. Bell, H. Yang. 2007. Impact of congestion charging on the transit market: An inter-modal equilibrium model. *Transportation Res.* **41A**(7) 703–713.
- Xiao, F., H. Yang, D. Han. 2007. Competition and efficiency of private toll roads. *Transportation Res.* **41B**(3) 292–308.
- Yang, H., H.-J. Huang. 2005. *Mathematical and Economic Theory of Road Pricing*. Elsevier, Oxford, UK.
- Yang, H., W. Xu, B. Heydecker. 2009. Bounding the efficiency of road pricing. *Transport. Res. Part E*. Forthcoming.