Efficiency Loss in Market Mechanisms for Resource Allocation

by

Ramesh Johari

A.B., Mathematics, Harvard University (1998)
Certificate of Advanced Study in Mathematics, University of Cambridge (1999)

Submitted to the
Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
in Electrical Engineering and Computer Science

at the

Massachusetts Institute of Technology

June 2004

Signature of Author . . . .
Department of Electrical Engineering and Computer Science
May 18, 2004

Certified by . . . .
John N. Tsitsiklis
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by . . . .
Arthur C. Smith
Chairman, Committee on Graduate Students
Department of Electrical Engineering and Computer Science

Efficiency Loss in Market Mechanisms for Resource Allocation

by

Ramesh Johari

Submitted to the Department of Electrical Engineering
and Computer Science on May 18, 2004,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

This thesis addresses a problem at the nexus of engineering, computer science, and economics: in large scale, decentralized systems, how can we efficiently allocate scarce resources among competing interests? On one hand, constraints are imposed on the system designer by the inherent architecture of any large scale system. These constraints are counterbalanced by the need to design mechanisms that efficiently allocate resources, even when the system is being used by participants who have only their own best interests at stake.

We consider the design of resource allocation mechanisms in such environments. The analytic approach we pursue is characterized by four salient features. First, the monetary value of resource allocation is measured by the aggregate surplus (aggregate utility less aggregate cost) achieved at a given allocation. An efficient allocation is one which maximizes aggregate surplus. Second, we focus on market-clearing mechanisms, which set a single price to ensure demand equals supply. Third, all the mechanisms we consider ensure a fully efficient allocation if market participants do not anticipate the effects of their actions on market-clearing prices. Finally, when market participants are price anticipating, full efficiency is generally not achieved, and we quantify the efficiency loss.

We make two main contributions. First, for three economic environments, we consider specific market mechanisms and exactly quantify the efficiency loss in these environments when market participants are price anticipating. The first two environments address settings where multiple consumers compete to acquire a share of a resource in either fixed or elastic supply; these models are motivated by resource allocation in communication networks. The third environment addresses competition between multiple producers to satisfy an inelastic demand; this model is motivated by market design in power systems.

Our second contribution is to establish that, under reasonable conditions, the mechanisms we consider minimize efficiency loss when market participants anticipate the effects of their actions on market-clearing prices. Formally, we show that in a class of market-clearing mechanisms satisfying certain simple mathematical assumptions and for which there exist fully efficient competitive equilibria, the mechanisms we consider uniquely minimize efficiency loss when market participants are price anticipating.

Thesis Supervisor: John N. Tsitsiklis
Title: Professor of Electrical Engineering

Acknowledgments

First and foremost, I am deeply indebted to my advisor, Professor John Tsitsiklis. His encouragement, support, and advice have been immensely valuable, both in personal and professional terms. I am particularly grateful for his emphasis on simplicity and elegance in research, and for the genuine concern he has shown for my development as an academic.

I am also grateful to Professor Frank Kelly for his enthusiasm and counsel, a constant throughout my graduate career. I first became familiar with network pricing problems in conversations with him, and this thesis bears the hallmark of lessons learned during my stay in Cambridge under his tutelage.

Special thanks go to my thesis committee members, Professor Abhijit Banerjee and Professor Robert Gallager. Each devoted significant time and effort to my thesis, and their suggestions and comments led to substantial improvement in the final product.

The Laboratory for Information and Decision Systems (LIDS) provided an ideal work environment for the interdisciplinary work of this thesis. I am particularly grateful to Professor Dimitri Bertsekas, for helping me navigate the subtleties of convex analysis; to Professor Vincent Chan, for his steady hand in directing the lab; and to Professor Sanjoy Mitter, for his intellectual stewardship of my time at MIT. I am in the debt of these faculty and many others at MIT for their support over the last four years.

I have also benefited from time spent with colleagues at MIT. I would especially like to thank Shie Mannor, whose dedication to research helped keep me disciplined; and Emin Martinian, a frequent counterpart for lunch conversations, both academic and otherwise. Finally, I was particularly fortunate to have had Constantine Caramanis as an officemate, with whom I've shared many interesting discussions since my days as an undergraduate.

Last, I owe a great debt to Hsin Chau, for her support, encouragement, and sense of humor; her companionship has been invaluable in all facets of life.

This research was supported by a National Science Foundation Graduate Research Fellowship, by the Defense Advanced Research Projects Agency under the Next Generation Internet Initiative, and by the Army Research Office under grant DAAD10-00-1-0466.

Preliminaries

Notation

We use $\mathbb{R}$ to denote the real numbers, and $\mathbb{R}^+$ to denote $[0, \infty)$. Italics will be used to denote scalars, e.g., x. Boldface will be used to denote vectors, e.g., $\mathbf{x} = (x_1, \ldots, x_n)$. When x is a scalar, the notation $(x)^+$ will be used to denote the positive part of x; i.e., $(x)^+ = x$ if $x \geq 0$, and $(x)^+ = 0$ if $x \leq 0$. If $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^m$, and $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, we will use $\mathbf{x}_{-i}$ to denote the components of $\mathbf{x}$ other than $\mathbf{x}_i$; that is, $\mathbf{x}_{-i} = (\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n)$. Throughout the thesis, if $f : (\mathbb{R}^m)^n \to \mathbb{R}$ is a real-valued function of n vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^m$, we let $f(\mathbf{x}_i; \mathbf{x}_{-i})$ denote the function f as a function of $\mathbf{x}_i$ while keeping the components $\mathbf{x}_{-i}$ fixed.

Convex analytic methods play a key role in this thesis, and we collect some required notions here [14, 103]. An *extended real-valued function* is a function $f : \mathbb{R}^n \to [-\infty, \infty]$; such a function is called *proper* if $f(\mathbf{x}) > -\infty$ for all $\mathbf{x}$, and $f(\mathbf{x}) < \infty$ for at least one $\mathbf{x}$. We say that a vector $\boldsymbol{\gamma} \in \mathbb{R}^n$ is a *subgradient* of an extended real-valued function f at $\mathbf{x}$ if for all $\overline{\mathbf{x}} \in \mathbb{R}^n$, we have:

$$f(\overline{\mathbf{x}}) \geq f(\mathbf{x}) + \boldsymbol{\gamma}^\top (\overline{\mathbf{x}} - \mathbf{x}).$$

The *subdifferential* of f at $\mathbf{x}$, denoted $\partial f(\mathbf{x})$, is the set of all subgradients of f at $\mathbf{x}$. We say that f is *subdifferentiable* at $\mathbf{x}$ if $\partial f(\mathbf{x}) \neq \emptyset$. We will typically be interested in subgradients of a convex function f, and *supergradients* of a concave function f. A vector $\boldsymbol{\gamma}$ is a supergradient of f if $-\boldsymbol{\gamma}$ is a subgradient of $-f$; thus we denote the *superdifferential* of f at $\mathbf{x}$ by $-\partial[-f(\mathbf{x})]$.

For extended real-valued functions $f : \mathbb{R} \to [-\infty, \infty]$, we will require some additional concepts. We denote the right directional derivative of f at x by $\partial^+ f(x)/\partial x$ and left directional derivative of f at x by $\partial^- f(x)/\partial x$ (if these exist). If f is convex, then $\partial f(x) = [\partial^- f(x)/\partial x, \partial^+ f(x)/\partial x]$, provided the directional derivatives exist.

Prerequisites

The main prerequisites for this thesis are a background in real analysis at the level of Rudin [110], as well as some facility with convex optimization and elementary convex analysis. Sources for background on convex optimization include the books by Whittle [145], Bertsekas [13], and Boyd and Vandenberghe [17], while background on convex analysis may be found in the texts by Bertsekas et al. [14] and Rockafellar [103].

Microeconomics (particularly market theory) and game theory also play a key role in this thesis, and some basic knowledge of the two fields is helpful. The text by Varian provides a concise introduction to microeconomic theory [137], while the textbook by Mas-Colell et al. provides deeper coverage [82]. As to game theory, in this thesis we will only use elementary concepts from game theory, particularly Nash equilibrium; however, some understanding of the modeling issues is helpful. For this purpose, see the books by Fudenberg and Tirole [43], Myerson [89], and Osborne and Rubinstein [96] (where the last reference is a concise introduction for the uninitiated reader).

Bibliographic Note

Portions of the content of Chapter 2 appear in the paper by Johari and Tsitsiklis [60]; exceptions are Sections 2.1.3, 2.3, 2.4.3, and 2.5.1. Sections 3.1, 3.2, 3.3, and 3.4 will appear in the paper by Johari et al. [58].

Contents

List of Figures

CHAPTER 1

Introduction

> *But man has almost constant occasion for the help of his brethren, and it is in vain for him to expect it from their benevolence only. He will be more likely to prevail if he can interest their self-love in his favour, and show them that it is for their own advantage to do for him what he requires of them. . . . It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest.*
>
> —Adam Smith, *The Wealth of Nations*, Book I, Chapter II [125]

This thesis addresses a problem at the nexus of engineering, computer science, and economics: in large scale, decentralized systems, how can we efficiently allocate scarce resources among competing interests? On one hand, constraints are imposed on the system designer by the inherent architecture of any large scale system. These constraints are counterbalanced by the need to design mechanisms that efficiently allocate resources, even when the system is being used by participants who have only their own best interests at stake.

Our inspiration is drawn primarily from communication networks and power systems. Communication networks, particularly the Internet, have a distributed structure which prohibits the implementation of sophisticated centralized mechanisms to allocate network resources among end users. On the other hand, the growth of the Internet has led to increasingly diverse traffic sharing the same network, making efficient allocation of network resources difficult to ensure. Power systems are also characterized by large scale, but by contrast, typically run a collection of markets that determine clearing prices for electricity at nodes throughout regional networks. In this setting demand is highly inelastic, largely because the variation in prices seen by most power consumers is on a slow timescale. In general, the chosen market designs for large scale power networks do not protect against the exercise of market power in the presence of highly inelastic demand.

Our approach in this thesis is to design mechanisms which take into account architectural features of these systems, while ensuring that efficient resource allocation is achieved. The analytic approach we have chosen is characterized by four salient features, described in detail in each of the next four sections. In Section 1.1, we describe our metric of efficiency, the *aggregate surplus* of consumers and producers. In Section 1.2, we restrict attention to *market-clearing mechanisms* which set a single price to "clear the market" between suppliers and consumers of resources. In Section 1.3, we note that all the mechanisms we consider achieve efficient allocations when market participants act as *price takers*. Finally, in Section 1.4, we outline our measurement of efficiency loss when market participants are *price anticipating*. In Section 1.5, we develop a detailed example that highlights each of the points discussed in Sections 1.1 to 1.4. We briefly outline the contributions of the thesis in Section 1.6.

■ 1.1 Consumers, Producers, and Aggregate Surplus

In this section we will describe the general resource allocation setting we will consider, and our definition of efficiency. All the resource allocation problems we describe consist of two types of market participants: *consumers* and *producers*. Consumers demand an allocation of a resource; and producers supply that resource. We identify each consumer r with a *utility function* $U_r(d_r)$, which defines the monetary value to consumer r of receiving d_r units of the scarce resource, where $d_r \geq 0$. Similarly, we identify each producer n with a *cost function* $C_n(s_n)$, which defines the monetary cost to producer n of supplying s_n units of the scarce resource, where $s_n \geq 0$. We will formalize detailed assumptions on these constructs throughout the thesis; but for the moment, we simply note that we will always assume both utility and cost are nondecreasing functions. The implication is that consumers value larger amounts of resources, while suppliers incur higher costs for producing larger amounts of resources.

A key assumption we have made is that both utility and cost are measured in monetary units. This assumption implies that there are actually two types of goods in the resource allocation settings we consider: the first is the scarce resource under consideration (data rate, electric power, etc.), and the second is money. Suppose then that a consumer with utility function U receives a resource allocation d, but makes a payment w; then the net *payoff* to this consumer is:

$$U(d) - w. \tag{1.1}$$

On the other hand, suppose that a producer with cost function C produces a supply s, but receives revenue w; then the net payoff to this producer is:

$$w - C(s). \tag{1.2}$$

Thus payoffs give the *net* monetary benefit to consumers and producers, taking into account both the money paid or received, and the resource allocation received or produced. The separable form of the payoffs we see here is a direct consequence of the fact that utility and cost are measured in monetary units. Environments where payoffs have this separable form are known as *quasilinear environments* [46, 82].

We are searching for mechanisms that achieve efficient allocation of resources. We adopt as our notion of efficiency the well known concept of *Pareto efficiency*: an allocation is Pareto efficient if the benefit to one market participant cannot be strictly increased without simultaneously strictly decreasing the benefit to another player. Throughout this thesis, we always use the term *efficient allocation* to refer to a Pareto efficient allocation.

We now characterize the implications of Pareto efficiency on our resource allocation model. For simplicity, we start by considering a model consisting of only two consumers with utility functions U_1 and U_2, and a single resource of *inelastic supply* S; that is, the maximum available supply is fixed at S units. We can interpret inelastic supply in terms of a producer with a discontinuous cost function: if the supply of the resource available is exactly S units, then it is *as if* a single producer supplies the resource, with cost function $C(s)$ given by:

$$C(s) = \begin{cases} 0, & \text{if } s \leq S; \\ \infty, & \text{if } s > S. \end{cases}$$

Thus the producer incurs zero cost if at most S units must be supplied, and infinite cost otherwise. In particular, in searching for a Pareto efficient allocation, the assumption of inelastic supply implies a *constraint* that the total allocation made to the two consumers cannot exceed the available supply S.

The key assumption we make is that the two consumers may freely exchange currency. We claim that under this circumstance, any Pareto efficient allocation $\mathbf{d} = (d_1, d_2)$ to the two consumers must be an optimal solution to the following optimization problem:

$$\begin{aligned} \text{maximize} \quad & U_1(d_1) + U_2(d_2) && (1.3) \\ \text{subject to} \quad & d_1 + d_2 \leq S; && (1.4) \\ & d_1, d_2 \geq 0. && (1.5) \end{aligned}$$

The intuition is clear: if a Pareto efficient allocation is not an optimal solution to (1.3)-(1.5), then there must exist a solution to (1.3)-(1.5) and a vector of monetary transfers between the consumers that collectively leave both consumers better off than the original allocation. We now demonstrate this fact formally. Suppose that (d_1, d_2) is Pareto efficient, but that for another feasible solution (d_1^*, d_2^*) to (1.3)-(1.5), we have:

$$U_1(d_1) + U_2(d_2) < U_1(d_1^*) + U_2(d_2^*).$$

We can assume without loss of generality that $U_1(d_1) < U_1(d_1^*)$. In this case, suppose the consumers shift to $\mathbf{d}^*$ from $\mathbf{d}$, but that in addition player 1 pays player 2 an amount

$w = (U_2(d_2) - U_2(d_2^*))^+$. Then the payoff to player 1 is now:

$$U_1(d_1^*) - w = U_1(d_1^*) - (U_2(d_2) - U_2(d_2^*))^+ > U_1(d_1),$$

while the payoff to player 2 becomes:

$$U_2(d_2^*) + w = U_2(d_2^*) + (U_2(d_2) - U_2(d_2^*))^+ \geq U_2(d_2).$$

Thus player 1 is strictly better off than before, and player 2 is no worse off than before; so $\mathbf{d}$ could not have been Pareto efficient.

The preceding simple story is in fact quite general: as long as we allow arbitrary monetary transfers between market participants, any Pareto efficient allocation must maximize aggregate utility less aggregate cost. This quantity is known as the *aggregate surplus*, or Marshallian aggregate surplus (after the economist Alfred Marshall, though an early precursor was considered by Dupuit; see [35, 81, 82, 134]). Aggregate surplus denotes the net monetary benefit to the economy under a chosen allocation. In this work we will consider three instances of the aggregate surplus maximization problem, which we refer to as the *SYSTEM* problem throughout the thesis. In Chapter 2, we consider a model where multiple consumers bid for a single resource in inelastic supply; in this case the *SYSTEM* problem reduces to maximization of *aggregate utility* subject to the supply constraint. In Chapter 3, we consider a model where multiple consumers bid for a single resource in elastic supply, and in this case the *SYSTEM* problem is written directly as maximization of aggregate utility less aggregate cost. Finally, in Chapter 4, we consider a model where multiple producers bid to satisfy an inelastic demand. In this case it is *as if* there exists a single consumer with utility $-\infty$ if less than the fixed demand is produced, and utility zero otherwise. Thus the *SYSTEM* problem reduces to minimization of *aggregate cost* subject to the demand constraint.

We note that in an engineering context, "efficient" allocation of resources is often simply interpreted as a requirement that all available resources be allocated, without any specification of the distribution of the allocation over players. In the example illustrated above, an efficient allocation would then be *any* allocation where the capacity constraint (1.4) holds with equality. Indeed, this is precisely Pareto efficiency if consumers have utilities that are strictly increasing in their allocation, and if no monetary transfers are available to the consumers.

However, as we have seen, measuring cost and utility in monetary units reduces welfare measurement to a simple and convenient quantity, the aggregate surplus. When we introduce money into the system, a Pareto efficient allocation must not only fully allocate available resources, but also maximize aggregate surplus. For the example above, a Pareto efficient allocation should not only satisfy the constraint (1.4), but also maximize the objective function (1.3). However, we note that the use of aggregate utility, and aggregate surplus more generally, as a welfare metric has traditionally been a point of great debate in both economics and philosophy. The heart of this debate can be traced to the fact that the very notion of aggregate utility presupposes the ability to *compare* the utilities of different members of society, even though such a comparison may not be possible. One might assume that incomparability of preferences is resolved by measuring all utilities and costs in monetary units; however, members of society from different income classes place different values on currency itself, and thus one cannot claim that "a dollar is a dollar is a dollar."

The notion of maximization of aggregate utility as a desirable goal for society was first proposed by Bentham, the father of utilitarianism [11]. Marshall advanced the application of quasilinear payoffs to market theory, and developed the resulting notion of maximization of aggregate surplus as a Pareto efficient allocation rule [81]. But the quasilinear payoff model was critiqued for its presumption that comparison of utilities of different market participants is possible; an illuminating discussion of these issues in utility theory is provided in the pair of surveys by Stigler [128, 129]. The assumption that utility was inherently incomparable eventually led to the famous impossibility theorem of Arrow in social choice theory [3, 117].

For the purposes of this thesis, we only advise the reader that our focus on quasilinear environments, though well motivated, is by no means canonical. The broader issue in social choice is that market participants may not be motivated by only their monetary payoff, and in addition, as a community they may not be interested in achieving a Pareto efficient allocation. Instead, fairness concerns may be paramount, such as ensuring that the *number* of consumers or producers who choose to participate in the marketplace is as large as possible. Such questions suggest interesting departures from the quasilinear payoff models used in this thesis. For further study, a thorough discussion of preferences and utility involving elements of both philosophy and economics may be found in the volume of Sen [118].

■ 1.2 Market-Clearing Mechanisms

Having defined the maximal aggregate surplus as the benchmark we hope to attain, we now consider the problem of defining resource allocation mechanisms which reach that goal. Our focus in this thesis will be on *market-clearing mechanisms*. These mechanisms choose a single price so that demand equals supply, i.e., to "clear the market."

Why consider single price market mechanisms, as opposed to price discriminatory solutions? First, our hope in this thesis is to advance the theory of resource allocation mechanisms which are feasible in large scale, distributed systems. In such systems, particularly the Internet, the fine-scale distinction of users needed to implement a sophisticated mechanism of price discrimination does not seem viable; for this reason, we are led to simpler pricing schemes. A mechanism which sets a single price has the advantage of *anonymity*: to determine the market-clearing price, the mechanism needs

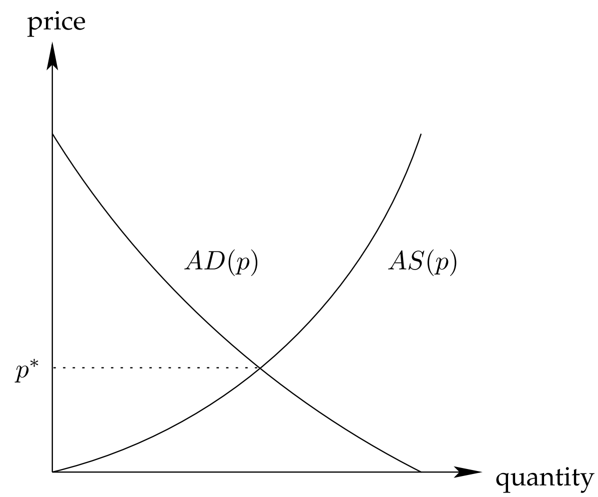

Figure 1-1. The market-clearing process: Each consumer r submits a demand function $D_r(p)$ to the market mechanism, and each producer n submits a supply function $S_n(p)$. These define the aggregate demand curve $AD(p) = \sum_r D_r(p)$, and the aggregate supply curve $AS(p) = \sum_n S_n(p)$. The price p^* is chosen so that supply equals demand, i.e., so that $AD(p^*) = AS(p^*)$. (Here and throughout the thesis, for market-clearing diagrams we adopt the standard convention from economics that quantity appears on the horizontal axis and price on the vertical axis.)

no knowledge of individual consumers and producers—only the aggregate supply and aggregate demand are required. (We will return to the discussion of scalability of efficient resource allocation mechanisms in the context of our review of mechanism design in the conclusion to the thesis, Chapter 6.) A second potential motivation for single price mechanisms comes from their use in practice, particularly in electricity markets; the "equity" of setting a single price for all market participants seems to have appeal from a social and political standpoint, and has led to widespread use of bidding systems which set a single price per node in an electricity grid [131].

We describe the basic operation of such a mechanism for the case of a single resource. Each consumer r chooses a demand function $D_r(p)$, which describes his demand for the resource as a function of the price p of that resource. Analogously, each producer n chooses a supply function $S_n(p)$, which describes the quantity the producer is willing to supply as a function of the price of the resource. The mechanism then chooses a single price p^* so that aggregate demand equals aggregate supply:

$$\sum_r D_r(p^*) = \sum_n S_n(p^*).$$

Each consumer r receives a resource allocation of $D_r(p^*)$, while each producer n produces a quantity $S_n(p^*)$. This process is graphically depicted in Figure 1-1.

In this thesis, we will generally be interested in either competition between consumers, or competition between producers; we leave open the analysis of models that simultaneously allow competition among both buyers and sellers. In Chapter 2, we consider a market-clearing mechanism where each consumer submits a demand function, and a price is chosen so that aggregate demand equals a preset inelastic supply. In Chapter 3, we consider a similar model, but where instead a price is chosen so that demand matches supply according to a preset elastic supply function. Finally, in Chapter 4, we consider a market-clearing mechanism where each producer submits a supply function, and a price is chosen so that aggregate supply equals a preset inelastic demand.

Models have previously been developed to understand market behavior when either suppliers submit supply functions, or consumers submit demand functions. In a seminal paper, Klemperer and Meyer characterized equilibria in markets where suppliers submit supply functions [69]. Competition among buyers who submit demand schedules was considered by Wilson [146]. Both these models allow market participants to submit nearly arbitrary supply or demand functions.

By contrast, we will consider mechanisms where consumers and producers are restricted to choose from *parametrized* demand or supply functions, where the parameter is a real scalar. For example, in Chapters 2 and 3, we allow a consumer r to choose a single scalar $w_r \geq 0$, and assume the resulting form of the demand function for that consumer is $D(p, w_r) = w_r/p$. The interpretation is that w_r is the total *willingness-to-pay* of consumer r, since regardless of the market-clearing price p^*, the payment made by consumer r will be $p^* D(p^*, w_r) = w_r$. (We investigate a simple example of such a mechanism in Section 1.5.)

We have two primary motivations for considering market mechanisms where participants submit parametrized demand or supply functions. First, in large scale decentralized systems such as modern communication networks, it seems unreasonable to expect the network to support transmission of arbitrary demand functions to widely dispersed resources running market-clearing processes. Instead, scalable market solutions for allocation of network resources must ensure the strategy space of users is "simple"—which we interpret here as a reduction in the dimension of the space of demand functions possible. (See also Kelly [63] for a similar argument regarding network pricing.)

A second motivation for the types of market mechanisms we consider is that reducing the strategy space of market participants might reduce inefficiency due to the exercise of market power. This is a point most forcefully made in the electricity markets, where currently firms may submit arbitrary supply functions, as proposed in the paper by Klemperer and Meyer [69]. In general, equilibria in supply functions may be highly inefficient when firms manipulate the market; and thus we are motivated to consider restrictions in the class of supply functions firms are allowed to submit, in hopes of improving the efficiency of the market mechanism. (A more detailed discussion of this issue may be found in the introduction to Chapter 4.)

1.3 Price Taking Behavior and Competitive Equilibrium

A central feature of all the market mechanisms we consider is that when market participants act as *price takers*, full efficiency can be achieved. Price takers are market participants who do not anticipate the effect of their strategic choices (i.e., their demand or supply function) on the eventual market-clearing price. An alternative situation is that market participants do in fact anticipate the effects of their actions on the market-clearing price—a situation we take up in the following section.

For simplicity, we will formalize price taking here only for a special case where multiple consumers compete for a single resource. In this case, if a consumer r has a utility function U_r and submits a demand function $D_r(p)$ to the market mechanism, the payoff to the consumer when the market-clearing price is p^* is given by:

$$U_r(D_r(p^*)) - p^* D_r(p^*).$$

The first term is the monetary value, or utility, to the consumer of the allocation $D_r(p^*)$; the second term is the payment made by the consumer when the price is p^* per unit consumed. Of course, the market-clearing price depends on the demand functions submitted by all consumers. In particular, as consumer r varies his demand function, the market-clearing price p^* will vary as well. Price taking behavior assumes this relationship is unknown to the consumers: all consumers take the price p^* as *fixed*, and then choose a demand function which optimizes their payoff given this fixed price p^*.

In this thesis, we will establish that for all the market mechanisms under consideration, price taking behavior leads to a Pareto efficient allocation. The formal statement of this proposition is that there exists a pair consisting of a price and a vector of strategies for all market participants such that: (1) the price is the market-clearing price given the composite strategy vector of the market participants; and (2) each of the participants has chosen their strategy to maximize their payoff given the fixed price. Such a pair is known as a *competitive equilibrium*. We will establish existence of competitive equilibria, and then establish that at competitive equilibria the resulting allocations maximize aggregate surplus; for the models of Chapters 2 and 3, these are results of Kelly [62] and Kelly et al. [65], respectively. Informally, these results state that a single price can be chosen so that individual optimization by market participants yields a globally efficient outcome.

The fact that competitive equilibria yield Pareto efficient allocations is a central result in market theory, the *first fundamental theorem of welfare economics* [82]. (The second fundamental theorem states that under sufficient assumptions, any Pareto efficient allocation can be achieved as a competitive equilibrium.) These fundamental theorems are the cornerstone results of *general equilibrium theory* [5, 31], first developed by Walras [142]; for this reason competitive equilibria are also referred to as *Walrasian equilibria*. Marshall discussed Pareto efficiency of competitive equilibria in the special case where utilities are separable [81, 82, 130, 137]; this line of development is often referred to as *partial equilibrium analysis*, because the assumption that utility is separable may be interpreted as a reflection of the fact that consumers spend a small fraction of their income on the goods under consideration, and that the prices of all other goods are held constant. In this case it is reasonable to assume that all members of society will value currency identically relative to the goods under consideration (see the discussion in Section 1.1, as well as Chapter 10 of [82]).

The first fundamental theorem provides a first justification for the attractiveness of market mechanisms. Assuming that no market participants anticipate the effects of the their actions, market mechanisms provide a simple and decentralized method to ensure efficient allocation of resources. But the assumption of limited price anticipation is quite a strong one, particularly if only a few market participants compete at any given time. In communication networks, one argument in favor of competitive equilibria is that the number of end users is enormous, and each user competes for only a small fraction of overall network resources. However, if we expect that pricing of network resources (such as transmission capacity) occurs only at high levels of aggregation, then only a few service providers may be competing with each other to acquire network resources—and in this case the exercise of market power becomes possible. Similarly, in electricity markets only a few firms typically compete at any given node of a regional electricity grid, calling into question the assumption that price taking behavior and competitive equilibria will result.

■ 1.4 Price Anticipating Behavior and Nash Equilibrium

Because we cannot guarantee that market participants will be price takers, we turn our attention to the possibility that they may *anticipate* the effect of their actions on market-clearing prices; in economic terminology, the market participants are said to have *market power*.[1] In this case, each participant views the market-clearing price as a function of the composite strategy vector of all market participants. Thus the competition between market participants who are price anticipating is a *game*: the payoff of a given player is directly expressed as a function of his own strategy, as well as the strategies of all other players.

[1]The term "market power" can be somewhat misleading, because some market participants may actually achieve a *lower* payoff when they are price anticipating instead of price taking. For this reason, we typically use the more precise phrase "price anticipating" in most of the formal development of the thesis.

We will study these games through their *Nash equilibria*. A Nash equilibrium [90] is a strategy vector from which no player has a unilateral incentive to deviate; that is, keeping the strategies of other players other than i fixed, the strategy chosen by player i maximizes his payoff. We make two observations regarding the choice of Nash equilibrium as our solution concept. First, of course, it is a *static* concept: that is, it gives no direction as to the *dynamic* process by which equilibrium is reached. Second, the Nash equilibrium solution concept assumes a great deal of knowledge on the part of the players of the game, and a key point of debate involves whether or not players possess sufficient information to implement the Nash equilibrium. In regards to the second point, a limited rebuttal may be offered by noting that for the market mechanisms considered in this thesis, a market participant generally needs information only on his own strategy and the price of the resource to determine whether his chosen strategy is payoff maximizing. However, both the objections raised here are of critical importance in market modeling and game theory in general, and we will take up further discussion on both topics in the conclusion to the thesis, Chapter 6.

While competitive equilibria are guaranteed to achieve full efficiency, Nash equilibria when market participants are price anticipating do not generally ensure full efficiency. However, the aggregate surplus at a Nash equilibrium still captures the net monetary benefit to the economy at the allocation that the equilibrium achieves. In this thesis, we will focus on the following basic question: *how inefficient are the allocations achieved when market participants are price anticipating, relative to Pareto efficient allocations?* Formally, this question is answered by considering the *ratio* of aggregate surplus achieved at a Nash equilibrium to the maximum possible aggregate surplus. Thus we investigate the percentage of monetary benefit lost to the economy because market participants are able to anticipate the effects of their actions.

The fact that Nash equilibria of a game may not achieve full efficiency has been well known in the economics literature [33]. The economist Pigou observed early on that there may be a gap between the optimal performance of a system and that achieved by selfish optimization [101]. Perhaps the simplest such example is the well known game of the Prisoners' Dilemma, where the Nash equilibrium chooses payoffs to the two players which are strictly lower than payoffs each player would obtain at another strategy vector [96]. This insight has been a central theme of the theory of industrial organization as well, particularly for oligopoly models; see Tirole [134] for a survey of these issues. In recognition of the effects of market power, heuristic measures are often used to determine the efficiency loss in an environment where some participants may be price anticipating. For example, the U.S. Department of Justice uses the *Herfindahl index*, the sum of the squares of percentage market shares of all firms in a market, as a measure of market concentration [135]. Informally, a Herfindahl index in excess of 1800 is interpreted as a sign that significant market power may be present in an industry, and by implication this situation is considered to yield high efficiency loss. (A formal investigation of the relationship between the Herfindahl index and efficiency is attempted by [25], and surveyed by Shapiro [119].)

A recent surge of research, primarily driven by the computer science community, has focused on quantifying efficiency loss for specific game environments. Most of the results have focused on network routing, starting with the initial work of Koutsoupias and Papadimitriou [70]. In that paper the authors introduce the notion of a "coordination ratio," which Papadimitriou later refers to as the "price of anarchy" [99]; this is precisely the ratio of a given performance metric at the Nash equilibrium of a game relative to the optimal value of that performance metric. Subsequent works on routing models include the papers by Mavronicolas and Spirakis [84]; Czumaj and Voecking [24]; Roughgarden and Tardos [106, 107, 108, 109]; Correa, Schulz, and Stier Moses [21, 113, 114]; and Perakis [100]. In addition to these works, other papers explore efficiency loss in network design problems [2, 32, 37], as well as a special class of submodular games including facility location games [138]. The key advance made by this research is the quantification of the loss of efficiency at Nash equilibria in specific game environments, and the goal of this thesis is to establish a quantitative understanding of efficiency loss in market mechanisms.

■ 1.5 An Example

In this section we will work through an example in detail to illustrate the concepts previously discussed in this chapter. Our purpose is to elucidate the meaning of the terms and assumptions used through a simple model, as preparation for the mathematically rigorous treatment offered in the remainder of the thesis. In working through this example, we will follow the same order of presentation of concepts as Sections 1.1 to 1.4.

We consider a model consisting of two consumers competing for a resource in inelastic supply, as discussed in Section 1.1; this is a special case of the environment considered in much greater detail in Chapter 2. We assume the inelastic supply is fixed at $S = 1$ unit. Furthermore, we assume that each consumer has a linear utility function: $U_r(d_r) = \alpha_r d_r$, $r = 1, 2$, where $\alpha_1 > \alpha_2 > 0$. We recall that as shown in Section 1.1, a Pareto efficient allocation must solve the following optimization problem:

$$\text{maximize} \quad \alpha_1 d_1 + \alpha_2 d_2 \tag{1.6}$$

$$\text{subject to} \quad d_1 + d_2 \leq 1; \tag{1.7}$$

$$d_1, d_2 \geq 0. \tag{1.8}$$

This problem is identical to (1.3)-(1.5), but where we have substituted for the utility functions of the two consumers in the objective function (1.3), and where we have set the inelastic supply S equal to one unit in (1.4).

Since we have assumed $\alpha_1 > \alpha_2$, the unique optimal solution $\mathbf{d}^*$ to (1.6)-(1.8) allocates the entire supply to consumer 1, so that $d_1^* = 1$, $d_2^* = 0$; this is therefore the unique Pareto efficient allocation. This yields aggregate utility $\alpha_1 d_1^* + \alpha_2 d_2^* = \alpha_1$. Thus, the maximum possible aggregate monetary benefit to the system consisting of two consumers and a single resource of unit supply is exactly equal to α_1.

In Section 1.5.1, we develop a market-clearing mechanism for allocation of this resource. In Section 1.5.2, we analyze the performance of the mechanism when the consumers are price takers; we will see that there exists a unique competitive equilibrium, and that the resulting allocation is Pareto efficient. In Section 1.5.3, we consider the possibility that consumers are price anticipating; we show there exists a unique Nash equilibrium, and note that it is not Pareto efficient. Finally, in Section 1.5.4, we quantify the efficiency loss when consumers are price anticipating by comparing the aggregate utility at the Nash equilibrium to the maximal aggregate utility (i.e., α_1).

■ 1.5.1 A Market-Clearing Mechanism

In this section we develop a market-clearing mechanism for allocation of the scarce resource. Suppose that the supplier of the resource, or *resource manager*, wishes to efficiently allocate the unit supply among the two consumers. We will analyze the following simple scheme, an analogue of the mechanism proposed by Kelly [62]:

1. Each consumer $r = 1, 2$ submits a total payment, or *bid*, w_r that the consumer is willing to make. The interpretation of w_r is that regardless of the market-clearing price, consumer r will always consume an amount of the resource which makes his payment exactly equal to w_r. Formally, if we denote the demand of consumer r at a price $p > 0$ by $D(p, w_r)$, then the equality that must be satisfied is:

$$pD(p, w_r) = w_r, \quad \text{for all } p > 0.$$

   In other words, we can interpret the bid w_r as submission of a *demand function* $D(p, w_r) = w_r/p$.

2. The resource manager chooses a price to "clear the market," i.e., so that the entire unit supply is allocated. Formally, the manager chooses a market-clearing price p^* so that:

$$D(p^*, w_1) + D(p^*, w_2) = 1.$$

   If we substitute $D(p, w) = w/p$, we find that the market-clearing price $p^* = p^*(\mathbf{w})$ is given by:

$$p^*(\mathbf{w}) = w_1 + w_2. \tag{1.9}$$

   (For technical simplicity, we ignore the boundary case where $w_1 + w_2 = 0$; such details are addressed in greater mathematical depth in Chapter 2.)

3. The allocation made to consumer r is now:

$$D(p^*(\mathbf{w}), w_r) = \frac{w_r}{w_1 + w_2},$$

while the payment made by consumer r is exactly $p^*(\mathbf{w})D(p^*(\mathbf{w}), w_r) = w_r$.

In examining the operation of this mechanism, we have defined it in terms of demand functions. However, because of the special structure of the demand functions, the eventual operation of the mechanism is actually quite simple: each consumer r pays an amount w_r, and receives a fraction of the resource *in proportion* to his payment. Note that the total revenue to the resource manager is equal to $w_1 + w_2$, the sum of the payments from the two consumers.

■ 1.5.2 Price Taking Consumers

In this section we will assume the consumers do not anticipate the effects of their actions on the market-clearing price. To understand this point concretely, consider the mechanism from the point of view of consumer 1. Consumer 1 wishes to choose his bid w_1 to maximize his payoff, defined in (1.1). There are two possibilities: either consumer 1 realizes that changing w_1 will change the market-clearing price; or consumer 1 does not *anticipate* this effect, and takes the market-clearing price as fixed when choosing an optimal bid w_1. In the latter case we say consumer 1 is a *price taker*. We analyze the price taking model in this section; we consider price anticipating consumers in the next section.

If consumer 1 assumes the market-clearing price stays fixed at p as w_1 varies, then he solves:

$$\max_{w_1 \geq 0} \left[\alpha_1 \cdot \frac{w_1}{p} - w_1 \right]. \tag{1.10}$$

To parse this expression, observe that when consumer 1 submits a bid w_1, if the market-clearing price is p he receives an allocation $D(p, w_1) = w_1/p$; this yields in turn the utility to consumer 1, $\alpha_1 w_1/p$. This is the first term in the expression (1.10). The second term is the payment w_1 made by consumer 1. Thus, (1.10) expresses the fact that consumer 1 chooses w_1 to maximize his payoff (cf. (1.1)) while taking the market-clearing price p as given and invariant. A symmetric expression holds for consumer 2:

$$\max_{w_2 \geq 0} \left[\alpha_2 \cdot \frac{w_2}{p} - w_2 \right]. \tag{1.11}$$

We are searching for a *competitive equilibrium*, as described in Section 1.3. In our setting, a competitive equilibrium is a vector $\mathbf{w} = (w_1, w_2)$ where each consumer has optimally chosen his bid, while taking the market-clearing price $p = p^*(\mathbf{w})$ as fixed. Formally, we want a pair (w_1, w_2) such that: (1) the market-clearing price is

$p = p^*(\mathbf{w}) = w_1 + w_2$; (2) the bid w_1 is an optimal solution to (1.10), given the price p; and (3) the bid w_2 is an optimal solution to (1.11), given the price p. Since $\alpha_1 > \alpha_2 > 0$, it is straightforward to establish that there exists a vector (w_1, w_2) satisfying these conditions, given by $w_1 = \alpha_1$, $w_2 = 0$. To see this, note that when $(w_1, w_2) = (\alpha_1, 0)$, the market-clearing price is $p = p^*(\mathbf{w}) = w_1 + w_2 = \alpha_1$. Thus the payoff to consumer 1 at a bid $\overline{w}_1$, given by $\alpha\overline{w}_1/p - \overline{w}_1$, is identically zero; in particular, $w_1 = \alpha_1$ is an optimal choice for consumer 1 given the price p. On the other hand, the payoff to consumer 2 at a bid $\overline{w}_2$ is $\alpha_2\overline{w}_2/p - \overline{w}_2$. Since $\alpha_2 < \alpha_1 = p$, we have $\alpha_2/p - 1 < 0$, so the unique optimal choice for user 2 is $w_2 = 0$.

Thus the pair $w_1 = \alpha_1$, $w_2 = 0$ is a competitive equilibrium, with market-clearing price $p^*(\mathbf{w}) = \alpha_1$. (In fact, it is possible to show that this is the unique competitive equilibrium.) Furthermore, observe that at this equilibrium, consumer 1 receives the entire resource: $D(p^*(\mathbf{w}), w_1) = 1$, while $D(p^*(\mathbf{w}), w_2) = 0$. In particular, in light of our previous discussion, the allocation at the competitive equilibrium is an optimal solution to (1.6)-(1.8). We conclude that the competitive equilibrium allocation is Pareto efficient—a special case of the *first fundamental theorem of welfare economics* (see Section 1.3).

■ 1.5.3 Price Anticipating Consumers

Now suppose that each consumer *anticipates* the effect of a change in his bid on the market-clearing price. Again, for simplicity, let us consider the mechanism from the point of view of consumer 1. Suppose that consumer 2 submits a bid of w_2. When consumer 1 submits a bid of w_1, the market-clearing price is $p^*(\mathbf{w}) = w_1 + w_2$, and the resulting allocation to consumer 1 is $D(p^*(\mathbf{w}), w_1) = w_1/(w_1 + w_2)$. Now suppose consumer 1 anticipates that the market-clearing price will change when w_1 changes; that is, rather than fixing the market-clearing price p and then optimally choosing w_1, as in (1.10), consumer 1 now takes into account the functional dependence of $p^*(\mathbf{w})$ on w_1. Thus, given w_2, consumer 1 chooses w_1 to solve:

$$\max_{w_1 \geq 0} \left[\alpha_1 \cdot \frac{w_1}{w_1 + w_2} - w_1 \right]. \tag{1.12}$$

Note that this payoff is identical to (1.10), except that we have replaced the allocation w_1/p with the term $w_1/(w_1 + w_2)$, reflecting the dependence of the market-clearing price on w_1. Symmetrically, given w_1, consumer 2 chooses w_2 to solve:

$$\max_{w_2 \geq 0} \left[\alpha_2 \cdot \frac{w_2}{w_1 + w_2} - w_2 \right]. \tag{1.13}$$

(In the discussion to follow, we ignore boundary effects where $w_1 = 0$ or $w_2 = 0$; again, such details are addressed in Chapter 2.)

Note that the payoff of each consumer is dependent on the choice made by the other consumer; thus the equations (1.12)-(1.13) define a *game*. We will search for a Nash equilibrium of this game, i.e., a vector (w_1, w_2) such that: (1) w_1 is an optimal solution to (1.12) given w_2; and (2) symmetrically, w_2 is an optimal solution to (1.13) given w_1. For technical simplicity, we search only for a Nash equilibrium such that $w_1 > 0$, $w_2 > 0$. Notice that given $w_2 > 0$, the payoff (1.12) is concave in w_1; and given $w_1 > 0$, the payoff (1.13) is concave in w_2. Thus if we differentiate (1.12) with respect to w_1, and (1.13) with respect to w_2, a Nash equilibrium is identified by the following two necessary and sufficient optimality conditions:

$$\alpha_1 \left(\frac{1}{w_1 + w_2} - \frac{w_1}{(w_1 + w_2)^2} \right) = 1;$$
$$\alpha_2 \left(\frac{1}{w_1 + w_2} - \frac{w_2}{(w_1 + w_2)^2} \right) = 1.$$

It is straightforward to check that these equations have a unique solution (w_1^{NE}, w_2^{NE}), given by:

$$w_1^{NE} = \frac{\alpha_1^2 \alpha_2}{(\alpha_1 + \alpha_2)^2}; \qquad w_2^{NE} = \frac{\alpha_1 \alpha_2^2}{(\alpha_1 + \alpha_2)^2}.$$

Thus the vector (w_1^{NE}, w_2^{NE}) is a Nash equilibrium.

We now characterize the market-clearing price, allocation, payoffs, and revenue at this Nash equilibrium. It is straightforward to check that the market-clearing price is:

$$p^*(\mathbf{w}^{NE}) = w_1^{NE} + w_2^{NE} = \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}.$$

Note that this is also the revenue to the resource manager at the Nash equilibrium. Recall that the revenue to the resource manager at the competitive equilibrium was $w_1 + w_2 = \alpha_1$. Since $\alpha_1 > \alpha_2$, we conclude the revenue to the resource manager is *lower* at the Nash equilibrium than at the competitive equilibrium. This result can be shown to hold more generally; see Corollary 2.3.

The allocation to consumer 1 at the Nash equilibrium, denoted d_1^{NE}, is:

$$d_1^{NE} = D(p^*(\mathbf{w}^{NE}), w_1^{NE}) = \frac{w_1^{NE}}{p^*(\mathbf{w}^{NE})} = \frac{\alpha_1}{\alpha_1 + \alpha_2}. \tag{1.14}$$

Symmetrically, the allocation d_2^{NE} to consumer 2 is:

$$d_2^{NE} = D(p^*(\mathbf{w}^{NE}), w_1^{NE}) = \frac{w_2^{NE}}{p^*(\mathbf{w}^{NE})} = \frac{\alpha_2}{\alpha_1 + \alpha_2}. \tag{1.15}$$

Thus the payoff to consumer 1 is:

$$\alpha_1 d_1^{NE} - w_1^{NE} = \frac{\alpha_1^3}{(\alpha_1 + \alpha_2)^2}.$$

A symmetric expression holds for consumer 2. In particular, note that now both payoffs are positive to the consumers, whereas both consumers had payoff equal to zero at the competitive equilibrium. While the payoffs to consumers rise in this case, it is not generally true that every consumer improves his payoff at a Nash equilibrium; indeed, examples exist where the payoffs of consumers can actually *fall* relative to a competitive equilibrium, despite the fact that they are price anticipating. The intuition for this is that the market power gained by a single price anticipating consumer may be undermined by the market power of all other consumers who are also price anticipating.

■ 1.5.4 Efficiency Loss

Recall that the unique Pareto efficient allocation, i.e., the unique optimal solution to (1.6)-(1.8), is given by $d_1^* = 1$, $d_2^* = 0$. Thus, at the Nash equilibrium, the allocation $\mathbf{d}^{NE}$ defined by (1.14)-(1.15) cannot be Pareto efficient in general. In this section we ask: how inefficient is the allocation at the Nash equilibrium?

To answer this question, we must consider the net monetary benefit of the Nash equilibrium, which is measured by the aggregate utility at the allocation $\mathbf{d}^{NE}$:

$$\alpha_1 d_1^{NE} + \alpha_2 d_2^{NE} = \frac{\alpha_1^2 + \alpha_2^2}{\alpha_1 + \alpha_2}.$$

On the other hand, the net monetary benefit at the unique Pareto efficient allocation (which is also the competitive equilibrium allocation) is given by:

$$\alpha_1 d_1^* + \alpha_2 d_2^* = \alpha_1.$$

To make a measurement of efficiency loss which is independent of the currency in which we measure monetary value, we consider the *ratio* of Nash equilibrium aggregate utility to maximal aggregate utility:

$$\frac{\alpha_1 d_1^{NE} + \alpha_2 d_2^{NE}}{\alpha_1 d_1^* + \alpha_2 d_2^*} = \frac{\alpha_1^2 + \alpha_2^2}{\alpha_1^2 + \alpha_1 \alpha_2}. \tag{1.16}$$

The right hand side thus gives the percentage of the maximal aggregate surplus which is achieved at a Nash equilibrium, and exactly quantifies the efficiency loss when consumers are price anticipating.

Recall that we had assumed $0 < \alpha_2 < \alpha_1$. It is not difficult to establish that for fixed

α_1, the minimum value of (1.16) over $0 < \alpha_2 < \alpha_1$ is achieved when $\alpha_2 = (\sqrt{2} - 1)\alpha_1$, and the resulting minimum value is $2\sqrt{2} - 2 \approx 0.83$. Thus the efficiency loss when consumers are price anticipating in this model is no more than approximately 17%. In fact, we will show in Chapter 2 that under much more general assumptions on the utility functions of the consumers, the efficiency loss is no more than 25%.

To summarize the example, we have considered a simple market-clearing mechanism to allocate a single resource of fixed, unit supply. Each consumer chooses a bid, which represents the total amount he is willing to pay; a price is then chosen to clear the market. When the consumers act as price takers, there exists a competitive equilibrium, and the resulting allocation is Pareto efficient. When consumers are price anticipating, there exists a Nash equilibrium, but the resulting allocation is not Pareto efficient. Nevertheless, the efficiency loss is never more than approximately 17% when consumers are price anticipating.

■ 1.6 Contributions of This Thesis

This thesis makes two main contributions. First, for three economic environments, we consider specific market mechanisms and exactly quantify the efficiency loss in these environments when market participants are price anticipating. These environments are described in turn in Chapters 2, 3, and 4. Second, we show in Chapter 5 that under reasonable conditions, the mechanisms we consider minimize efficiency loss when market participants are price anticipating.

In Chapter 2, we consider a setting of multiple consumers and inelastic supply. We investigate a network resource allocation mechanism proposed by Kelly [62] where network users submit demand functions of the form $D(p, w) = w/p$, and a price is chosen so that aggregate demand is equal to the inelastic supply. As in Section 1.5, for the case of a single resource, this allocation mechanism allocates fractions of the resource to the users in *proportion* to their bids w. We establish that when users are price anticipating, aggregate utility falls by no more than 25% relative to the maximum possible. We also develop an extension of this result to a network setting, where users submit individual bids to each link in the network where they require service; we then discuss the implications of this model for the network pricing proposal made by Kelly in [62].

In Chapter 3, we consider the same class of demand functions as Chapter 2, but now consider a setting where supply is elastic; this is the model considered by Kelly et al. [65]. For this setting we establish that when users are price anticipating, aggregate surplus falls by no more than approximately 34% relative to the maximum possible; and as in Chapter 2, we extend this result to a network setting. In this chapter we also characterize efficiency loss for *Cournot games* [23], where users submit demand functions that are constant in p (i.e., $D(p) = d$ for all p). While efficiency loss is generally arbitrarily high for this mechanism, we establish that in several special cases of interest the efficiency loss when users are price anticipating is no more than $1/3$.

In Chapter 4, motivated by power systems, we consider a setting where multiple producers bid to satisfy an inelastic demand D. We consider a market mechanism where producers submit supply functions of the form $S(p, w) = D - w/p$, and a market-clearing price is chosen to ensure that aggregate supply is equal to the inelastic demand. In this case we interpret w as the portion of the total revenue foregone by a firm, since for any price p, the revenue to a firm bidding w is $pD - w$, while pD is the revenue to all firms. We establish that when producers are price anticipating, aggregate production cost rises by no more than a factor $1 + 1/(N - 2)$ relative to the minimum possible production cost, where $N > 2$ is the number of firms competing.

Chapters 2, 3, and 4 establish efficiency loss results for specific market mechanisms. Our second primary contribution in this thesis is the development of two results in Chapter 5 which characterize the mechanisms studied in Chapters 2 and 4 as the "best" choice of mechanism under reasonable assumptions. Formally, we show that in a class of market-clearing mechanisms satisfying certain simple mathematical assumptions and for which there exist fully efficient competitive equilibria, the mechanisms we consider in Chapters 2 and 4 uniquely minimize efficiency loss when market participants are price anticipating. These results justify the attention devoted to understanding the market mechanisms studied in Chapters 2 and 4; furthermore, they clearly delineate conditions which must be violated if we hope to achieve higher efficiency guarantees than those provided by the results of Chapters 2 and 4.

Each of the chapter introductions discusses the content of that chapter, so we do not survey results in detail here. In particular, the "Chapter Outline" at the start of each chapter gives a roadmap to the results of that chapter. We recommend that Chapter 2 be read prior to Chapter 3. Chapter 4 is largely independent of the material in Chapters 2 and 3. Finally, a survey of Chapters 2 and 4 would be helpful in reading Chapter 5. The conclusion to the thesis, Chapter 6, discusses open issues and questions raised by the thesis, particularly related to dynamics and mechanism design in distributed environments.

CHAPTER 2

Multiple Consumers, Inelastic Supply

The current Internet is used by a widely heterogeneous population of users; not only are different types of traffic sharing the same network, but different end users place different values on their perceived network performance. As a result, characterizing "good" use of the network is difficult: how should resources be shared between a file transfer and a peer-to-peer connection? Partly in response to this heterogeneity, a variety of models for *congestion pricing* in the future Internet have emerged. These models propose a traditional economic solution to the problem of heterogeneous demand: they treat the collection of network resources as a market, and price their use accordingly.

The last decade has witnessed a dramatic rise in research suggesting the use of market mechanisms to manage congestion in networks, starting with research on congestion pricing in ATM networks, and continuing to subsequent efforts to develop congestion pricing for the Internet. See, for example, the critique by Shenker et al. [122] for an early overview of some of the issues involved; the book by Songhurst [126] for a review of work related to congestion charging in ATM networks; and the papers of Falkner [38] and Briscoe et al. [18] for more recent discussion.

Perhaps the simplest method of network pricing is a simple flat rate approach [94]: each user of a network pays a fixed fee for network use, independent of the actual resources consumed. The problem with such a simple scheme, of course, is that resources may not be allocated efficiently among the competing heterogeneous demands. Despite simple proposals to extend the flat rate pricing paradigm to include a few service classes [93], in general flat rate pricing is inefficient—see, for example, [77]. Such opinions have recently been supported through the experimental work of the INDEX project [36].

In response to the basic inefficiency of flat rate pricing, *usage based pricing* gained support. Under such a scheme, users are charged based on the impact their traffic has on network performance. The basic goal of such a method is to provide better feedback on congestion to users through a price signal; the hope is that the users will respond to such signals and achieve an efficient network operating point. Because the Internet is a large scale, decentralized system, any such scheme is forced to make a tradeoff: there is a compromise between sophistication of a pricing scheme and the ability to implement it in a distributed manner [51].

One of the earlier, more prominent proposals for congestion pricing is the "smart market" of MacKie-Mason and Varian [78]. Their approach uses an auction for data rate at each network resource: each user submits a bid for the total amount they are willing to pay for network service, and the network delivers available service to the highest bidders. Using classical results on Vickrey-Clarke-Groves (VCG) mechanisms [20, 50, 139], MacKie-Mason and Varian show that under their auction mechanism, each network user will have an incentive to truthfully reveal their valuation for network resources, and that the resulting allocation of network resources will be efficient (in the sense that resources will be allocated to those users who value them most highly).

Despite the appealing simplicity of the auction mechanism proposed by MacKie-Mason and Varian, a key objection is that the mechanism is not scalable in a distributed setting, since each network resource must perform a complex computation to determine the users which receive service. Indeed, the information requirements of implementing a VCG mechanism in a network can be quite high [116]. For this reason, later proposals by Lazar and Semret [76, 116] and Shu and Varaiya [123] attempt to devise distributed auction mechanisms. (We will return to this connection in our discussion of mechanism design for network resource allocation problems; see Section 6.3.)

Rather than attempting to simplify the implementation of an auction for network resources, we adopt an alternative approach in this chapter. The mechanism we consider views network resource allocation as a market-clearing process: users submit demand functions expressing their desire for network resources as a function of the prices of those resources, and the network chooses prices to ensure aggregate demand is exactly equal to available supply. The market mechanism is constrained by two features: first, the demand functions chosen by users must have a "simple" description, since they are to be communicated across a network; and second, the computation of prices of network resources should not require central coordination.

The specific framework we investigate was first proposed in the seminal work of Kelly [62]. Kelly considers a model consisting of a single network manager who wishes to allocate network capacity efficiently among a collection of users, each endowed with a utility function depending on their allocated rate. In [62], a market is proposed where each user submits a "bid," or willingness-to-pay per unit time, to the network; the network accepts these submitted bids and determines the price of each network link. A user is then allocated rate in proportion to his bid, and inversely proportional to the price of links he wishes to use. Under certain assumptions, it is shown in [62] that such a scheme maximizes aggregate utility.

In the special case where the network consists of only a single link, a user is allocated a fraction of the link equal to his bid divided by the sum of all users' bids. This "proportional" allocation mechanism has been considered in a variety of other contexts as well. Subsequent to Kelly's work, La and Anantharam suggested a means by which the proportional allocation mechanism might be implemented in a network employing window-based congestion control, such as the Internet [73, 74]. Hajek and Gopalakrishnan have considered a proportional allocation mechanism in the context of Internet autonomous system competition [52]. They suggest that smaller Internet providers might bid for resources from larger Internet providers upstream using the proportional allocation mechanism. In the economics literature, such a mechanism has been referred to as a "raffle"; it has been analyzed in the context of charitable fundraising [34]. In the computer science community, this mechanism is known as the "proportional share" mechanism, where it has been investigated for time-sharing of resources [132].

In this chapter, we wish to understand the extent to which the analysis proposed in [62] accurately models the interactions of network users. Specifically, a fundamental assumption in the model of [62] is that each user acts as a price taker; that is, users do not anticipate the effect of their actions on the prices of the links. In contrast, we relax this assumption, and ask whether price anticipating behavior significantly worsens the performance of the network. Such a relaxation is motivated by the fact that a large enough user may be able to elicit the exact response of network prices to changes in his strategy; see [22] for a model of an intelligent software agent that might mimic this task. If we assume that users can anticipate the effects of their actions, then the model becomes a game, and we will show that the Nash equilibria of this game lead to allocations at which the aggregate utility is no worse than 75% of the maximal aggregate utility.

Chapter Outline

The remainder of the chapter is organized as follows. In Section 2.1 we give background on the model formulation. We recapitulate the key results of [62], and precisely define the notion of price taking and competitive equilibrium. We prove the main theorem of [62] for a single link: if users are price taking, then aggregate utility is maximized. We then consider a game where users are price anticipating. We give a proof of a result due to Hajek and Gopalakrishnan establishing existence and uniqueness of a Nash equilibrium, by showing that at a Nash equilibrium, it is as if aggregate utility is maximized but with *modified* utility functions [52]. (Less general forms of this result have been previously established in the literature; see Section 2.1.2 for details.) We then establish several corollaries in Section 2.1.3. We first show that revenues to the link manager may be arbitrarily low at the Nash equilibrium relative to the competitive equilibrium; nevertheless, we show that if all utility functions are linear, then revenues to the link manager at a Nash equilibrium are no less than 50% of the revenues obtained if the link manager were to use a Vickrey auction.

In Section 2.2, we consider the loss of efficiency at the Nash equilibrium of the single link game. Theorem 2.6 is a key result of this chapter: when users are price anticipating, the efficiency loss is no more than 25%. The key insight is that the worst case occurs when utility functions are linear. We use this fact to explicitly construct the worst case game in the proof of Theorem 2.6. In addition to this result, we show in Corollary 2.8 that in an appropriate limit where each user consumes a negligible fraction of the available link rate, the ratio of Nash equilibrium aggregate utility to the maximal aggregate utility converges to one. Such a result is a "competitive limit" [82], demonstrating that as the number of users becomes large, if no user is a "large" consumer then it is *as if* all users are price takers.

In Section 2.3, we consider a model consisting of multiple profit maximizing network providers. In particular, we investigate whether network providers will have an incentive to truthfully reveal their link capacity. For a simple model of parallel links, where each link is controlled by an independent network provider, we show that when users have linear utility functions the unique profit maximizing strategy for each provider is to truthfully declare their capacity. A consequence of this result is that the loss of efficiency is no more than 25% when users are price anticipating *and* link managers are profit maximizing, provided the utility functions of the users are linear.

In Section 2.4, we extend the analysis of Section 2.2 to networks. We consider a game where each user requests service from multiple links by submitting a bid to each link. Users have multiple routes available to them for sending traffic, so that this is a model including alternative routing. Links then allocate rates using the same scheme as in the single link model, and each user sends the maximum rate possible, given the vector of rates allocated from links in the network. Although this definition of the game is natural, we demonstrate that Nash equilibria may not exist, due to a discontinuity in the payoff functions of individual players. (This problem also arises in the single link setting, but is irrelevant there as long as at least two players share the link.) To address the discontinuity, we extend the strategy space by allowing each user to request a nonzero rate without submitting a positive bid to a link, if the total payment made by other users at that link is zero; this extension is sufficient to guarantee existence of a Nash equilibrium. Furthermore, if a Nash equilibrium exists in the original game, it corresponds naturally to a Nash equilibrium of the extended game. Finally, we show that in this network setting, the total utility achieved at any Nash equilibrium of the game is no less than 3/4 of the maximum possible aggregate utility. This extends the efficiency loss result from the single link case to the network setting.

The model of Section 2.4 is not identical to the original network pricing proposal of Kelly [62]. In particular, the proposal made in [62] requires each user to submit a *single bid* to the network, rather than individual bids to each link. In Section 2.4.3, we explore the consequences of this difference and the relationship between the network pricing model of Section 2.4.1 and the original proposal of [62].

In Section 2.5, we consider two extensions to the basic model of this chapter. First, in Section 2.5.1, we consider the possibility that link capacity is stochastic, rather than deterministic. We establish that such a model is identical to the model of Sections 2.1 and 2.2, for an appropriate choice of utility functions and with capacity equal to 1. This identification allows us to carry over the key results of the chapter to the setting where capacity is stochastic.

If we interpret the model proposed in [62] in a broader economic context, then a key feature is that the supplies of resources available are inelastic—that is, they do not vary with prices. Thus we may view the model of this chapter as a solution for a general class of resource allocation problems with inelastic supply. This motivates the resource allocation game in Section 2.5.2. We suppose that users bid for multiple resources, as in Section 2.4; but we no longer define utility as a function of the maximum rate that a user can send. Rather, we allow the user's utility to be *any* concave function of the vector of resources allocated. As an example of such a game, each resource may be a raw material, and each end user may be a manufacturing facility that takes these raw materials as input. Building on Section 2.5.1, we also allow the capacity of each resource to be randomly determined. We show that such a game can be analyzed using the same methods as Section 2.4, and in particular prove once again that the efficiency loss is no worse than 25% relative to the optimal aggregate utility.

■ 2.1 Preliminaries

Suppose R users share a communication link of capacity $C > 0$. Let d_r denote the rate allocated to user r. We assume that user r receives a *utility* equal to $U_r(d_r)$ if the allocated rate is d_r; we assume that utility is measured in monetary units. We make the following assumptions on the utility function.

Assumption 2.1
For each r, over the domain $d_r \geq 0$ the utility function $U_r(d_r)$ is concave, strictly increasing, and continuous; and over the domain $d_r > 0$, $U_r(d_r)$ is continuously differentiable. Furthermore, the right directional derivative at 0, denoted $U_r'(0)$, is finite.

We note that we make rather strong differentiability assumptions here on the utility functions; these assumptions are primarily made to ease the presentation. In Section 2.4, we will relax the assumption that U_r is differentiable.

Concavity in Assumption 2.1 corresponds to the assumption of *elastic* traffic, as defined by Shenker [121]. This is in fact quite a strong assumption in the setting of communication networks; note that elastic traffic typically refers to file transfers, while traffic such as telephone calls and video streams (with minimum rate requirements) may be modeled using nonconcave utility functions. For example, if a telephone call requires a minimum data rate D, the corresponding utility function might be zero for any rate less than D, and constant but positive for any rate greater than or equal to D. Indeed, an important research direction concerns development of resource allocation models for settings where users may have nonconcave utility as a function of the rate received.

Given complete knowledge and centralized control of the system, a natural problem for the network manager to try to solve is the following optimization problem [62]:

SYSTEM:

$$\text{maximize} \quad \sum_r U_r(d_r) \tag{2.1}$$

$$\text{subject to} \quad \sum_r d_r \leq C; \tag{2.2}$$

$$d_r \geq 0, \quad r = 1, \ldots, R. \tag{2.3}$$

Note that the objective function of this problem is the *aggregate utility*. This is the appropriate adaptation of the notion of aggregate surplus to a setting where supply is inelastic; see Section 1.1. Since the objective function is continuous and the feasible region is compact, an optimal solution $\mathbf{d} = (d_1, \ldots, d_R)$ exists. If the functions U_r are strictly concave, then the optimal solution is unique, since the feasible region is convex.

In general, the utility functions are not available to the link manager. As a result, we consider the following pricing scheme for rate allocation. Each user r gives a payment (also called a *bid*) of w_r to the link manager; we assume $w_r \geq 0$. Given the vector $\mathbf{w} = (w_1, \ldots, w_r)$, the link manager chooses a rate allocation $\mathbf{d} = (d_1, \ldots, d_r)$. We assume the manager treats all users alike—in other words, the network manager does not *price discriminate*. Each user is charged the same price $\mu > 0$, leading to $d_r = w_r/\mu$. We further assume the manager always seeks to allocate the entire link capacity C; in this case, following the analysis of [62], we expect the price μ to satisfy:

$$\sum_r \frac{w_r}{\mu} = C.$$

The preceding equality can only be satisfied if $\sum_r w_r > 0$, in which case we have:

$$\mu = \frac{\sum_r w_r}{C}. \tag{2.4}$$

In other words, if the manager chooses to allocate the entire available rate at the link, and does not price discriminate between users, then for every nonzero $\mathbf{w}$ there is a *unique* price $\mu > 0$ which must be chosen by the network, given by the previous equa-

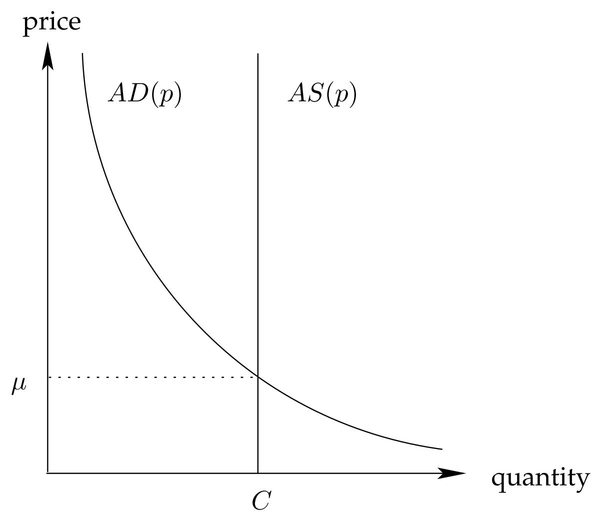

Figure 2-1. The market-clearing process with inelastic supply: Each consumer r chooses a willingness-to-pay w_r, which maps to a demand function $D(p, w_r) = w_r/p$. This defines the aggregate demand function $AD(p) = \sum_r D(p, w_r) = \sum_r w_r/p$. The aggregate supply function is $AS(p) = C$ for all p. The price μ is chosen so that supply equals demand, i.e., so that $\sum_r w_r/\mu = AD(\mu) = AS(\mu) = C$.

tion.

We can interpret this mechanism as a *market-clearing* process by which a price is set so that demand equals supply. To see this interpretation, note that when a user chooses a total payment w_r, it is as if the user has chosen a *demand function* $D(p, w_r) = w_r/p$ for $p > 0$. The demand function describes the amount of rate the user demands at any given price $p > 0$. The link manager then chooses a price μ so that $\sum_r D(\mu, w_r) = C$, i.e., so that the aggregate demand equals the supply C; see Figure 2-1. For the specific form of demand functions we consider here, this leads to the expression for μ given in (2.4). User r then receives a rate allocation given by $D(\mu, w_r)$, and makes a payment $\mu D(\mu, w_r) = w_r$. This interpretation of the mechanism we consider here will be further explored in Chapter 5, where we consider other market-clearing mechanisms for allocating a single resource in inelastic supply, with the users choosing demand functions from a family parametrized by a single scalar.

We note here that the interpretation of the mechanism in terms of users submitting demand functions bears strong resemblance to the work of Wilson [146] on auctions of divisible goods. Wilson considered a model where users submit demand functions, and a price is chosen to ensure that supply equals demand; this model has later been studied in the context of Treasury auctions [143]. The key difference between Wilson's model and the model of this chapter is the fact that the demand functions we consider are parametrized by a single scalar. As discussed in the introduction to the chapter, this decision is made so that the strategy space of the users is simple: rather than communicating an entire demand function across the network, the users only submit their total willingness-to-pay. (The interpretation of Wilson's model as a "demand function equilibrium" bears close relation to the "supply function equilibrium" studied by Klemperer and Meyer [69], which we discuss in Chapter 4 in the context of electricity markets.)

In the remainder of the section, we consider two different models for how users might interact with this price mechanism. In Section 2.1.1, we consider a model where users do not anticipate the effect of their bids on the price, and establish existence of a competitive equilibrium (a result due to Kelly [62]). Furthermore, this competitive equilibrium leads to an allocation which is an optimal solution to *SYSTEM*. In Section 2.1.2, we change the model and assume users are price anticipating, and establish existence and uniqueness of a Nash equilibrium (a result due to Hajek and Gopalakrishnan [52]). Section 2.2 then considers the loss of efficiency at this Nash equilibrium, relative to the optimal solution to *SYSTEM*.

■ 2.1.1 Price Taking Users and Competitive Equilibrium

In this section, we consider a *competitive equilibrium* between the users and the link manager [82], following the development of Kelly [62]. A central assumption in the definition of competitive equilibrium is that each user does not anticipate the effect of their payment w_r on the price μ, i.e., each user acts as a *price taker*. In this case, given a price $\mu > 0$, user r acts to maximize the following payoff function over $w_r \geq 0$:

$$P_r(w_r; \mu) = U_r\left(\frac{w_r}{\mu}\right) - w_r. \tag{2.5}$$

The first term represents the utility to user r of receiving a rate allocation equal to w_r/μ; the second term is the payment w_r made to the manager. Observe that since utility is measured in monetary units, the payoff is *quasilinear* in money [82].

We now say a pair $(\mathbf{w}, \mu)$ with $\mathbf{w} \geq 0$ and $\mu > 0$ is a *competitive equilibrium* if users maximize their payoff as defined in (2.5), and the network "clears the market" by setting the price μ according to (2.4):

$$P_r(w_r; \mu) \geq P_r(\overline{w}_r; \mu) \quad \text{for } \overline{w}_r \geq 0, \;\; r = 1, \ldots, R; \tag{2.6}$$

$$\mu = \frac{\sum_r w_r}{C}. \tag{2.7}$$

Kelly shows in [62] that when users are price takers, there exists a competitive equilibrium, and the resulting allocation is an optimal solution to *SYSTEM*. This is formalized in the following theorem, adapted from [62]; we also present a proof for completeness.

Theorem 2.1 (Kelly, [62])
Suppose that Assumption 2.1 holds. Then there exists a competitive equilibrium, i.e., a vector

$\mathbf{w} = (w_1, \ldots, w_R) \geq 0$ *and a scalar* $\mu > 0$ *satisfying* (2.6)-(2.7).

In this case, the scalar μ *is uniquely determined, and the vector* $\mathbf{d} = \mathbf{w}/\mu$ *is an optimal solution to SYSTEM. If the functions* U_r *are strictly concave, then* $\mathbf{w}$ *is uniquely determined as well.*

Proof. The key idea in the proof is to use Lagrangian techniques to establish that optimality conditions for (2.6)-(2.7) are identical to the optimality conditions for the problem *SYSTEM*, under the identification $\mathbf{d} = \mathbf{w}/\mu$.

Step 1: Given $\mu > 0$, $\mathbf{w}$ *satisfies* (2.6) *if and only if:*

$$U_r'\left(\frac{w_r}{\mu}\right) = \mu, \quad \text{if } w_r > 0; \tag{2.8}$$

$$U_r'(0) \leq \mu, \quad \text{if } w_r = 0. \tag{2.9}$$

Indeed, since U_r is concave, P_r is concave as well; and thus (2.8)-(2.9) are necessary and sufficient optimality conditions for (2.6).

Step 2: There exists a vector $\mathbf{d} \geq 0$ *and a unique scalar* $\mu > 0$ *such that:*

$$U_r'(d_r) = \mu, \quad \text{if } d_r > 0; \tag{2.10}$$

$$U_r'(0) \leq \mu, \quad \text{if } d_r = 0; \tag{2.11}$$

$$\sum_r d_r = C. \tag{2.12}$$

The vector $\mathbf{d}$ *is then an optimal solution to SYSTEM. If the functions* U_r *are strictly concave, then* $\mathbf{d}$ *is unique as well.* Note that at least one optimal solution to *SYSTEM* exists since the feasible region is compact and the objective function is continuous. We form the Lagrangian for the problem *SYSTEM*:

$$\mathcal{L}(\mathbf{d}, \mu) = \sum_r U_r(d_r) - \mu\left(\sum_r d_r - C\right)$$

Here the second term is a penalty for the link capacity constraint. The Slater constraint qualification ([13], Section 5.3) holds for the problem *SYSTEM* at the point $\mathbf{d} = 0$, since then $0 = \sum_r d_r < C$; this guarantees the existence of a *Lagrange multiplier* μ. In other words, since the objective function is concave and the feasible region is convex, a feasible vector $\mathbf{d}$ is optimal if and only if there exists $\mu \geq 0$ such that the conditions (2.10)-(2.12) hold. Since there exists at least one optimal solution $\mathbf{d}$ to *SYSTEM*, there exists at least one pair $(\mathbf{d}, \mu)$ satisfying (2.10)-(2.12).

Since $C > 0$, at least one d_r is positive, so $\mu > 0$ (since U_r is strictly increasing). We now claim that μ is uniquely determined. Suppose not; then there exist $(\mathbf{d}, \mu)$, $(\overline{\mathbf{d}}, \overline{\mu})$

that satisfy (2.10)-(2.12), where (without loss of generality) $\mu < \overline{\mu}$. For any r such that $\overline{d}_r > 0$, we will have $U_r'(d_r) \leq \mu < \overline{\mu} = U_r'(\overline{d}_r)$, which implies that $d_r > \overline{d}_r > 0$. Summing over all r, we obtain $\sum_r d_r > \sum_r \overline{d}_r$, which contradicts the feasibility condition $\sum_r d_r = C = \sum_r \overline{d}_r$. Thus μ is unique.

Step 3: If the pair $(\mathbf{d}, \mu)$ *satisfies* (2.10)-(2.12), *and we let* $\mathbf{w} = \mu\mathbf{d}$, *then the pair* $(\mathbf{w}, \mu)$ *satisfies* (2.6)-(2.7). By Step 2, $\mu > 0$; thus, under the identification $\mathbf{w} = \mu\mathbf{d}$, (2.12) becomes equivalent to (2.7). Furthermore, (2.10)-(2.11) become equivalent to (2.8)-(2.9); by Step 1, this guarantees that (2.6) holds.

Step 4: If $\mathbf{w}$ *and* $\mu > 0$ *satisfy* (2.6)-(2.7), *and we let* $\mathbf{d} = \mathbf{w}/\mu$, *then the pair* $(\mathbf{d}, \mu)$ *satisfies* (2.10)-(2.12). We simply reverse the argument of Step 3. Under the identification $\mathbf{d} = \mathbf{w}/\mu$, (2.8)-(2.9) become equivalent to (2.10)-(2.11); and (2.7) becomes equivalent to (2.12).

Step 5: Completing the proof. By Steps 2 and 3, there exists a vector $\mathbf{w}$ and a scalar $\mu > 0$ satisfying (2.6)-(2.7); by Step 4, μ is uniquely determined, and the vector $\mathbf{d} = \mathbf{w}/\mu$ is an optimal solution to *SYSTEM*. Finally, if the utility functions U_r are strictly concave, then by Steps 2 and 4, $\mathbf{w}$ is uniquely determined as well (since the transformation from $(\mathbf{w}, \mu)$ to $(\mathbf{d}, \mu)$ is one-to-one). □

Theorem 2.1 shows that under the assumption that the users of the link behave as price takers, there exists a bid vector $\mathbf{w}$ where all users have optimally chosen their bids w_r, with respect to the given price $\mu = \sum_r w_r/C$; and at this "equilibrium," aggregate utility is maximized. However, when the price taking assumption is violated, the model changes into a game and the guarantee of Theorem 2.1 is no longer valid. We investigate this game in the following section.

■ 2.1.2 Price Anticipating Users and Nash Equilibrium

We now consider an alternative model where the users of a single link are price anticipating, rather than price takers. The key difference is that while the payoff function P_r takes the price μ as a fixed parameter in (2.5), price anticipating users will realize that μ is set according to (2.4), and adjust their payoff accordingly; this makes the model a game between the R players.

We use the notation $\mathbf{w}_{-r}$ to denote the vector of all bids by users other than r; i.e., $\mathbf{w}_{-r} = (w_1, w_2, \ldots, w_{r-1}, w_{r+1}, \ldots, w_R)$. Given $\mathbf{w}_{-r}$, each user r chooses w_r to maximize:

$$Q_r(w_r; \mathbf{w}_{-r}) = \begin{cases} U_r\left(\dfrac{w_r}{\sum_s w_s} C\right) - w_r, & \text{if } w_r > 0; \\ \\ U_r(0), & \text{if } w_r = 0. \end{cases} \tag{2.13}$$

over nonnegative w_r. The second condition is required so that the rate allocation to user r is zero when $w_r = 0$, even if all other users choose $\mathbf{w}_{-r}$ so that $\sum_{s \neq r} w_s = 0$. The payoff function Q_r is similar to the payoff function P_r, except that the user anticipates that the network will set the price μ according to (2.4). A *Nash equilibrium* of the game defined by $(Q_1, \ldots, Q_R)$ is a vector $\mathbf{w} \geq 0$ such that for all r:

$$Q_r(w_r; \mathbf{w}_{-r}) \geq Q_r(\overline{w}_r; \mathbf{w}_{-r}), \quad \text{for all } \overline{w}_r \geq 0. \tag{2.14}$$

Note that the payoff function in (2.13) may be discontinuous at $w_r = 0$, if $\sum_{s \neq r} w_s = 0$. This discontinuity may preclude existence of a Nash equilibrium, as the following example shows.

Example 2.1
Suppose there is a single user with strictly increasing utility function U. In this case, the user is not playing a game against anyone else, so any positive payment results in the entire link being allocated to the single user. The payoff to the user is thus:

$$Q(w) = \begin{cases} U(C) - w, & \text{if } w > 0; \\ U(0), & \text{if } w = 0. \end{cases}$$

Since U has been assumed to be strictly increasing, we know $U(C) > U(0)$. Thus, at a bid of $w = 0$, a profitable deviation for the user is any bid $\overline{w}$ such that $0 < \overline{w} < U(C) - U(0)$. On the other hand, at any bid $w > 0$, a profitable deviation for the user is any bid $\overline{w}$ such that $0 < \overline{w} < w$. Thus no optimal choice of bid exists for the user, which implies that no Nash equilibrium exists. □

We will find the previous discontinuity plays a larger role in the network context, where an extended strategy space is required to ensure existence of a Nash equilibrium. In the single link setting, Hajek and Gopalakrishnan have shown that there exists a unique Nash equilibrium when multiple users share the link, by showing that at a Nash equilibrium it is *as if* the users are solving another optimization problem of the same form as the problem *SYSTEM*, but with "modified" utility functions. This is formalized in the following theorem, adapted from [52]; we also present a proof for completeness.

Theorem 2.2 (Hajek and Gopalakrishnan, [52])
Suppose that $R > 1$, and that Assumption 2.1 holds. Then there exists a unique Nash equilibrium $\mathbf{w} \geq 0$ of the game defined by $(Q_1, \ldots, Q_R)$, and it satisfies $\sum_r w_r > 0$.

In this case, the vector $\mathbf{d}$ *defined by:*

$$d_r = \frac{w_r}{\sum_s w_s} C, \quad r = 1, \dots, R, \tag{2.15}$$

is the unique optimal solution to the following optimization problem:

GAME:

$$\text{maximize} \quad \sum_r \hat{U}_r(d_r) \tag{2.16}$$

$$\text{subject to} \quad \sum_r d_r \le C; \tag{2.17}$$

$$d_r \ge 0, \quad r = 1, \dots, R, \tag{2.18}$$

where

$$\hat{U}_r(d_r) = \left(1 - \frac{d_r}{C}\right) U_r(d_r) + \left(\frac{d_r}{C}\right) \left(\frac{1}{d_r} \int_0^{d_r} U_r(z)\, dz\right). \tag{2.19}$$

Proof. The proof proceeds in a number of steps. We first show that at a Nash equilibrium, at least two components of $\mathbf{w}$ must be positive. This suffices to show that the payoff function Q_r is strictly concave and continuously differentiable for each user r. We then establish necessary and sufficient conditions for $\mathbf{w}$ to be a Nash equilibrium; these conditions look similar to the optimality conditions (2.8)-(2.9) in the proof of Theorem 2.1, but for "modified" utility functions defined according to (2.19). Mirroring the proof of Theorem 2.1, we then show the correspondence between these conditions and the optimality conditions for the problem *GAME*. This correspondence establishes existence and uniqueness of a Nash equilibrium.

Step 1: If $\mathbf{w}$ *is a Nash equilibrium, then at least two coordinates of* $\mathbf{w}$ *are positive.* Fix a user r, and suppose $w_s = 0$ for every $s \ne r$. If $w_r > 0$, user r can maintain the same rate allocation and reduce his payment by reducing w_r slightly; and since U_r is strictly increasing, if $w_r = 0$, then user r can profitably deviate by infinitesimally increasing his bid w_r and capturing the entire link capacity C. Thus at a Nash equilibrium, $w_s > 0$ for some $s \ne r$. Since this holds for every user r, at least two coordinates of $\mathbf{w}$ must be positive.

Step 2: If the vector $\mathbf{w} \ge 0$ *has at least two positive components, then the function* $Q_r(\overline{w}_r; \mathbf{w}_{-r})$ *is strictly concave and continuously differentiable in* $\overline{w}_r$*, for* $\overline{w}_r \ge 0$*.* This follows from (2.13), because when $\sum_{s \ne r} w_s > 0$, the expression $\overline{w}_r / (\overline{w}_r + \sum_{s \ne r} w_s)$ is a strictly increasing function of $\overline{w}_r$ (for $\overline{w}_r \ge 0$); in addition, $U_r(\cdot)$ is a strictly increasing, concave, and differentiable function by assumption.

Step 3: The vector $\mathbf{w}$ *is a Nash equilibrium if and only if at least two components of* $\mathbf{w}$ *are positive, and for each* r*, the following conditions hold:*

$$U_r'\left(\frac{w_r}{\sum_s w_s}C\right)\left(1-\frac{w_r}{\sum_s w_s}\right) = \frac{\sum_s w_s}{C}, \quad \text{if } w_r > 0; \tag{2.20}$$

$$U_r'(0) \leq \frac{\sum_s w_s}{C}, \quad \text{if } w_r = 0. \tag{2.21}$$

Let $\mathbf{w}$ be a Nash equilibrium. By Steps 1 and 2, $\mathbf{w}$ has at least two positive components and $Q_r(\overline{w}_r; \mathbf{w}_{-r})$ is strictly concave and continuously differentiable in $\overline{w}_r \geq 0$. Thus w_r must be the unique maximizer of $Q_r(\overline{w}_r; \mathbf{w}_{-r})$ over $\overline{w}_r \geq 0$, and satisfy the following first order optimality conditions:

$$\frac{\partial Q_r}{\partial w_r}(w_r; \mathbf{w}_{-r}) \begin{cases} = 0, & \text{if } w_r > 0; \\ \leq 0, & \text{if } w_r = 0. \end{cases}$$

After multiplying through by $\sum_s w_s/C$, these are precisely the conditions (2.20)-(2.21).

Conversely, suppose that $\mathbf{w}$ has at least two strictly positive components, and satisfies (2.20)-(2.21). Then we may simply reverse the argument: by Step 2, $Q_r(\overline{w}_r; \mathbf{w}_{-r})$ is strictly concave and continuously differentiable in $\overline{w}_r \geq 0$, and in this case the conditions (2.20)-(2.21) imply that w_r maximizes $Q_r(\overline{w}_r; \mathbf{w}_{-r})$ over $\overline{w}_r \geq 0$. Thus $\mathbf{w}$ is a Nash equilibrium.

If we let $\mu = \sum_r w_r/C$, note that the conditions (2.20)-(2.21) have the same form as the optimality conditions (2.8)-(2.9), but for a different utility function given by $\hat{U}_r$. We now use this relationship to complete the proof in a manner similar to the proof of Theorem 2.1.

Step 4: The function $\hat{U}_r$ *defined in* (2.19) *is strictly concave and strictly increasing over* $0 \leq d_r \leq C$. The proof follows by differentiating, which yields $\hat{U}_r'(d_r) = U_r'(d_r)(1 - d_r/C)$. Since U_r is concave and strictly increasing, we know that $U_r'(d_r) > 0$, and that U_r' is nonincreasing; we conclude that $\hat{U}_r'(d_r)$ is nonnegative and strictly decreasing in d_r over the region $0 \leq d_r \leq C$, as required.

Step 5: There exists a unique vector $\mathbf{d}$ *and scalar* ρ *such that:*

$$U_r'(d_r)\left(1-\frac{d_r}{C}\right) = \rho, \quad \text{if } d_r > 0; \tag{2.22}$$

$$U_r'(0) \leq \rho, \quad \text{if } d_r = 0; \tag{2.23}$$

$$\sum_r d_r = C. \tag{2.24}$$

The vector $\mathbf{d}$ *is then the unique optimal solution to GAME.* By Step 4, since $\hat{U}_r$ is continuous and strictly concave over the convex, compact feasible region for each r, we know that *GAME* has a unique optimal solution. This optimal solution $\mathbf{d}$ is uniquely identified by the stationarity conditions (2.22)-(2.23), together with the constraint that $\sum_r d_r \leq C$. Since $\hat{U}_r$ is strictly increasing for each r, the constraint (2.24) must hold as well. That ρ is unique then follows because at least one d_r must be strictly positive at the unique optimal solution to *GAME*.

Step 6: If $(\mathbf{d}, \rho)$ *satisfies* (2.22)-(2.24)*, then the vector* $\mathbf{w} = \rho\mathbf{d}$ *is a Nash equilibrium.* We first check that at least two components of $\mathbf{d}$ are positive, and that $\rho > 0$. We know from (2.24) that at least one component of $\mathbf{d}$ is strictly positive. Suppose now that $d_r > 0$, and $d_s = 0$ for $s \neq r$. Then we must have $d_r = C$. But then by (2.22), we have $\rho = 0$; on the other hand, since U_s is strictly increasing and concave, we have $U_s'(0) > 0$ for all s, so (2.23) cannot hold for $s \neq r$. Thus at least two components of $\mathbf{d}$ are positive. In this case, it is seen from (2.22) that $\rho > 0$ as well.

By Step 3, to check that $\mathbf{w} = \rho\mathbf{d}$ is a Nash equilibrium, we must only check the stationarity conditions (2.20)-(2.21). We simply note that under the identification $\mathbf{w} = \rho\mathbf{d}$, using (2.24) we have that:

$$\rho = \frac{\sum_r w_r}{C}; \text{ and } d_r = \frac{w_r}{\sum_s w_s} C.$$

Substitution of these expressions into (2.22)-(2.23) leads immediately to (2.20)-(2.21). Thus $\mathbf{w}$ is a Nash equilibrium.

Step 7: If $\mathbf{w}$ *is a Nash equilibrium, then the vector* $\mathbf{d}$ *defined by* (2.15) *and scalar* ρ *defined by* $\rho = (\sum_r w_r)/C$ *are the unique solution to* (2.22)-(2.24)*.* We simply reverse the argument of Step 6. By Step 3, $\mathbf{w}$ satisfies (2.20)-(2.21). Under the identifications of (2.15) and $\rho = \sum_r w_r/C$, we find that $\mathbf{d}$ and ρ satisfy (2.22)-(2.24). By Step 5, such a pair $(\mathbf{d}, \rho)$ is unique.

Step 8: There exists a unique Nash equilibrium $\mathbf{w}$*, and the vector* $\mathbf{d}$ *defined by* (2.15) *is the unique optimal solution to GAME.* This conclusion is now straightforward. Existence follows by Steps 5 and 6, and uniqueness follows by Step 7 (since the transformation from $\mathbf{w}$ to $(\mathbf{d}, \rho)$ is one-to-one). Finally, that $\mathbf{d}$ is an optimal solution to *GAME* follows by Steps 5 and 7. □

Theorem 2.2 shows that the unique Nash equilibrium of the single link game is characterized by the optimization problem *GAME*. Other games have also profited from such relationships—notably traffic routing games, in which Nash equilibria can be found as solutions to a global optimization problem. Roughgarden and Tardos use this fact to their advantage in computing efficiency loss for such games [108]; Correa, Schulz, and Stier Moses also use this relationship to consider routing games in capacitated networks [21].

Theorem 2.2 is also closely related to *potential games* [86, 105], where best responses of players are characterized by changes in a global potential function. In such games, the global minima of the potential function correspond to Nash equilibria, as we observed for the problem *GAME*. However, it can be shown that despite this correspondence the objective function of the problem *GAME* is not a potential function.[1]

Finally, we note that for the game presented here, several authors have derived results similar to Theorem 2.2. Gibbens and Kelly [45] considered the special case where all the functions U_r are linear, and demonstrated existence and uniqueness of the Nash equilibrium in this setting. The first result for general utility functions was given by La and Anantharam [73], who showed that if the users' strategies are restricted to a compact set $[W_{\min}, W_{\max}]$, where $0 < W_{\min} < W_{\max} < \infty$, then there exists a unique Nash equilibrium. Maheswaran and Basar consider a model where a fixed value of $\varepsilon > 0$ is added to the price of the link [79]; the allocation to user r is thus $d_r = w_r/(\sum_s w_s + \varepsilon)$, which avoids the possible discontinuity of Q_r when $w_r = 0$. The authors demonstrate existence and uniqueness of the Nash equilibrium in this setting. It is possible to use the model of [79] to show existence (but not uniqueness) of the Nash equilibrium of the game defined by $(Q_1, \dots, Q_R)$, by taking a limit as $\varepsilon \to 0$; indeed, such a limit forms the basis of our proof of existence of Nash equilibria in the network context (see Theorem 2.12).

■ 2.1.3 Corollaries

The ability of users to anticipate their effect on prices is a form of *market power* in economic terminology [82]. One effect of market power is typically a loss of efficiency; in our model, this loss is no more than 25% of the optimal aggregate utility, as summarized by Theorem 2.6. Another effect, however, is that as the users gain market power, the resource manager loses market power; formally, the revenues to the resource manager drop when users anticipate their effect on prices. This is summarized in the following theorem.

Corollary 2.3
Suppose that $R > 1$, and that Assumption 2.1 holds. Let $(\mathbf{w}^S, \mu)$ be a competitive equilibrium, and let $\rho = \sum_r w_r/C$, where $\mathbf{w}$ is the unique Nash equilibrium of Theorem 2.2. Then $0 \leq \rho < \mu$.

[1] To see why, we first note that for (2.16) to yield a potential function, it must be the case that the derivative of (2.16) with respect to the strategy w_r of user r must have the same sign as the derivative of $Q_r(w_r; \mathbf{w}_{-r})$ with respect to w_r, for fixed $\mathbf{w}_{-r}$. On the other hand, observe that (2.16) is a function of $\mathbf{d}$, not $\mathbf{w}$; if we substitute (2.15) and differentiate with respect to w_r, the resulting derivative need not have the same sign as the derivative of $Q_r(w_r; \mathbf{w}_{-r})$ with respect to w_r.

Proof. The fact that $\rho \geq 0$ follows from the definition. Let $\mathbf{d}^S$ be any optimal solution to *SYSTEM*, and let $\mathbf{d}^G$ be the unique optimal solution to *GAME*. Then we know:

$$\begin{aligned} U_r'(d_r^S) &= \mu, \quad \text{if } d_r^S > 0; \\ &\leq \mu, \quad \text{if } d_r^S = 0; \\ \hat{U}_r'(d_r^G) = \left(1 - \frac{d_r^G}{C}\right) U_r'(d_r^G) &= \rho, \quad \text{if } d_r^G > 0; \\ &\leq \rho, \quad \text{if } d_r^G = 0. \end{aligned}$$

These statements follow from the proofs of Theorems 2.1 and 2.2; recall that both ρ and μ are uniquely determined.

Suppose that $\rho \geq \mu$. Consider any r such that $d_r^S = 0$. Then $U_r'(0) \leq \mu$, so $U_r'(0) \leq \rho$; thus $d_r^G = 0$ as well. Thus if $d_r^G > 0$, then $d_r^S > 0$.

Consider any r such that $d_r^G > 0$; we know $d_r^G < C$, since at least two components of $\mathbf{d}^G$ are strictly positive (by the proof of Theorem 2.2). Then:

$$U_r'(d_r^G) = \rho \left(1 - \frac{d_r^G}{C}\right)^{-1} > \rho \geq \mu.$$

On the other hand, since $d_r^G > 0$ implies $d_r^S > 0$, we know that $U_r'(d_r^S) = \mu$. So we must have $d_r^G < d_r^S$. Since this is true for all r where $d_r^G > 0$, we conclude that $C = \sum_r d_r^G < \sum_r d_r^S = C$, a contradiction. Thus we must have had $\rho < \mu$. □

The following corollary is immediate from the preceding proof.

Corollary 2.4
Suppose that $R > 1$, and that Assumption 2.1 holds. Let $A^S \subseteq \{1, \dots, R\}$ be the set of users with positive rate at any optimal solution $\mathbf{d}^S$ of SYSTEM, and let $A^G \subseteq \{1, \dots, R\}$ be the set of users with positive rate at the unique optimal solution $\mathbf{d}^G$ of GAME. Then $A^S \subseteq A^G$.

Recall that the resource manager's revenue when users are not price anticipating is given by $\sum_r w_r = \mu C$ (from (2.7)), where $(\mathbf{w}, \mu)$ satisfies the conditions of Theorem 2.1. Similarly, when users play the game $(Q_1, \dots, Q_R)$, the resource manager's revenue is $\sum_r w_r = \rho C$, where $\mathbf{w}$ is the unique Nash equilibrium of Theorem 2.2. Corollary 2.3 thus shows that the resource manager's revenue is *strictly lower* when users anticipate the effect of their actions on the link price.

Both the lower and upper bounds on ρ in Corollary 2.3 are essentially tight. We first give an example where ρ is arbitrarily close to μ. Consider a system with R users, where $U_r(d_r) = d_r$ for all r. Then $\mu = 1$, since $U_r'(d_r) = 1$ for all r. At the unique Nash equilibrium of the game, $d_r = 1/R$ for all r, and $\rho = 1 - d_r = 1 - 1/R$. Thus, as $R \to \infty$, we find that $\rho \to \mu$.

To see that ρ may become arbitrarily small, consider two users sharing a link of capacity 1, where $U_1(d_1) = d_1$, and $U_2(d_2) = \varepsilon d_2$. Then at the efficient allocation, $d_1 = C, d_2 = 0$, and the unique competitive equilibrium price is $\mu = 1$. On the other hand, at the Nash equilibrium we have:

$$1 - d_1 = \varepsilon(1 - d_2) = \rho,$$

and $d_1 + d_2 = 1$. Combining these relations, we find that:

$$\rho = \frac{\varepsilon}{1+\varepsilon}.$$

As $\varepsilon \to 0$, the revenues ρ to the network manager approach zero as well; note that in this case the ratio ρ/μ also goes to zero, so the *percentage* of revenues lost can be arbitrarily high.

However, at least one positive result is possible, in the case where all utilities are linear. Let $U_r(d_r) = \alpha_r d_r$, with $R > 1$. Assume without loss of generality that $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_R > 0$. The following corollary shows that the ratio of the Nash equilibrium price ρ to the second highest slope α_2 is no lower than $1/2$. Recall that under a Vickrey auction with linear utilities [139], the revenue to the auctioneer is given by the second highest valuation for the commodity, given by $\alpha_2 C$. Thus the following corollary guarantees that the revenue to the resource manager is no worse than 50% of the revenue under a Vickrey auction.

Corollary 2.5
Suppose that $R > 1$, and that for each r, $U_r(d_r) = \alpha_r d_r$, where $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_R > 0$. Let $\rho = \sum_r w_r/C$, where $\mathbf{w}$ is the unique Nash equilibrium of Theorem 2.2. Then $\rho \geq \alpha_2/2$, and this bound is tight.

Proof. Let $\mathbf{d}^G$ denote the allocation at the Nash equilibrium. From Theorem 2.2, we know that at least two users have $d_r^G > 0$ at the Nash equilibrium. Furthermore, from the structure of the modified utility function (2.19), it is clear that if $\alpha_r \geq \alpha_s$, then $d_r^G \geq d_s^G$. We conclude that $d_1^G \geq d_2^G$, and:

$$\rho = \alpha_1 \left(1 - \frac{d_1^G}{C}\right) = \alpha_2 \left(1 - \frac{d_2^G}{C}\right).$$

Now suppose that $\rho < \alpha_2/2$. Then from the preceding relation, this is only possible if $d_2^G > C/2$. Since $d_1^G \geq d_2^G$, we have $d_1^G > C/2$. But then $d_1^G + d_2^G > C$, which violates the capacity constraint in *GAME*. Thus we must have $\rho \geq \alpha_2/2$. The bound holds with equality if $R = 2$ and $\alpha_1 = \alpha_2 > 0$. □

2.2 Efficiency Loss: The Single Link Case

We let $\mathbf{d}^S$ denote an optimal solution to *SYSTEM*, and let $\mathbf{d}^G$ denote the unique optimal solution to *GAME*. We now investigate the efficiency loss of this system; that is, how much utility is lost because the users are price anticipating? To answer this question, we must compare the utility $\sum_r U_r(d_r^G)$ obtained when the users fully evaluate the effect of their actions on the price, and the utility $\sum_r U_r(d_r^S)$ obtained by choosing the point which maximizes aggregate utility. (We know, of course, that $\sum_r U_r(d_r^G) \leq \sum_r U_r(d_r^S)$, by definition of $\mathbf{d}^S$.)

An easy lower bound on $\sum_r \hat{U}_r(d_r^G)$ may be constructed by using the modified utility functions $\hat{U}_r$ defined in (2.19). Notice that $\hat{U}_r(d_r)$ may be viewed as the "expectation" of U_r with respect to a probability distribution which places a mass of $1 - d_r/C$ on the rate d_r (the first term of (2.19)), and uniformly distributes the remaining mass of d_r/C on the interval $[0, d_r]$ (the second term of (2.19)). We now use this interpretation to show that the efficiency loss is no more than 50% when users are price anticipating.

Assume that $U_r(0) \geq 0$. Since U_r is strictly increasing, we know that:

$$\frac{1}{d_r} \int_0^{d_r} U_r(z)\, dz \leq U_r(d_r).$$

From expression (2.19), we may conclude that for $0 \leq d_r \leq C$, we have $\hat{U}_r(d_r) \leq U_r(d_r)$.

By concavity of U_r, we have the following inequality for $0 \leq z \leq d_r$:

$$U_r(z) \geq \frac{z}{d_r} U_r(d_r) + \left(1 - \frac{z}{d_r}\right) U_r(0).$$

Now since $U_r(0) \geq 0$, this inequality reduces to:

$$U_r(z) \geq \frac{z}{d_r} U_r(d_r).$$

Integrating both sides from 0 to d_r, we have:

$$\int_0^{d_r} U_r(z)\, dz \geq \frac{d_r}{2} U_r(d_r).$$

We have the trivial bound that $U_r(d_r) \geq \frac{1}{2} U_r(d_r)$, since U_r is strictly increasing and

$U_r(0) \geq 0$. Thus:

$$\begin{aligned}\hat{U}_r(d_r) &= \left(1 - \frac{d_r}{C}\right) U_r(d_r) + \left(\frac{d_r}{C}\right)\left(\frac{1}{d_r}\int_0^{d_r} U_r(z)\, dz\right) \\ &\geq \left(1 - \frac{d_r}{C}\right)\left(\frac{1}{2}U_r(d_r)\right) + \left(\frac{d_r}{C}\right)\left(\frac{1}{2}U_r(d_r)\right) = \frac{1}{2}U_r(d_r).\end{aligned}$$

Let $\mathbf{d}^S$ be any optimal solution of *SYSTEM*. Then:

$$\frac{1}{2}\sum_r U_r(d_r^S) \leq \sum_r \hat{U}_r(d_r^S).$$

Let $\mathbf{d}^G$ be the unique optimal solution to *GAME*; then since $\mathbf{d}^S$ is also feasible for *GAME*, we know that

$$\sum_r \hat{U}_r(d_r^S) \leq \sum_r \hat{U}_r(d_r^G).$$

Finally, we know that $\hat{U}_r(d_r^G) \leq U_r(d_r^G)$ for each r. Combining these inequalities yields:

$$\frac{1}{2}\sum_r U_r(d_r^S) \leq \sum_r \hat{U}_r(d_r^S) \leq \sum_r \hat{U}_r(d_r^G) \leq \sum_r U_r(d_r^G).$$

The preceding argument shows that the efficiency loss is no more than 50% when users are price anticipating. However, this bound is not tight. As we show in the following theorem, the efficiency loss is exactly 25% in the worst case.

Theorem 2.6

Suppose that $R > 1$, and that Assumption 2.1 holds. Suppose also that $U_r(0) \geq 0$ for all r. If $\mathbf{d}^S$ is any optimal solution to SYSTEM, and $\mathbf{d}^G$ is the unique optimal solution to GAME, then:

$$\sum_r U_r(d_r^G) \geq \frac{3}{4}\sum_r U_r(d_r^S).$$

Furthermore, this bound is tight: for every $\varepsilon > 0$, there exists a choice of R, and a choice of (linear) utility functions U_r, $r = 1, \ldots, R$, such that:

$$\sum_r U_r(d_r^G) \leq \left(\frac{3}{4} + \varepsilon\right)\left(\sum_r U_r(d_r^S)\right).$$

Proof. We first show that because of the assumption that U_r is concave and strictly increasing for each r, the worst case occurs with linear utility functions. This is summarized in the following lemma.

Lemma 2.7 *Suppose that Assumption 2.1 holds. Suppose that $U_r(0) \geq 0$ for all r. Let $\overline{\mathbf{d}} = (\overline{d}_1, \ldots, \overline{d}_r)$ satisfy $\sum_r \overline{d}_r \leq C$, and let $\mathbf{d}^S$ be any optimal solution to SYSTEM. Then the following inequality holds:*

$$\frac{\sum_r U_r(\overline{d}_r)}{\sum_r U_r(d_r^S)} \geq \frac{\sum_r U_r'(\overline{d}_r)\overline{d}_r}{\left(\max_r U_r'(\overline{d}_r)\right) C}. \tag{2.25}$$

Proof of Lemma. Using concavity, we have $U_r(d_r^S) \leq U_r(\overline{d}_r) + U_r'(\overline{d}_r)(d_r^S - \overline{d}_r)$; see Figure 2-2. Furthermore, since U_r is strictly increasing and nonnegative for each r, we have $\sum_r U_r(d_r^S) > 0$. Thus:

$$\frac{\sum_r U_r(\overline{d}_r)}{\sum_r U_r(d_r^S)} \geq \frac{\sum_r (U_r(\overline{d}_r) - U_r'(\overline{d}_r)\overline{d}_r) + \sum_r U_r'(\overline{d}_r)\overline{d}_r}{\sum_r (U_r(\overline{d}_r) - U_r'(\overline{d}_r)\overline{d}_r) + \sum_r U_r'(\overline{d}_r) d_r^S}.$$

Furthermore, since $\sum_r d_r^S = C$, we have the following trivial inequality:

$$\sum_r U_r'(\overline{d}_r) d_r^S \leq \left(\max_r U_r'(\overline{d}_r)\right) C.$$

This yields:

$$\frac{\sum_r U_r(\overline{d}_r)}{\sum_r U_r(d_r^S)} \geq \frac{\sum_r (U_r(\overline{d}_r) - U_r'(\overline{d}_r)\overline{d}_r) + \sum_r U_r'(\overline{d}_r)\overline{d}_r}{\sum_r (U_r(\overline{d}_r) - U_r'(\overline{d}_r)\overline{d}_r) + \left(\max_r U_r'(\overline{d}_r)\right) C}.$$

Now notice that because we have assumed $U_r(0) \geq 0$, we again have by concavity that $U_r'(\overline{d}_r)\overline{d}_r \leq U_r(\overline{d}_r)$. Thus the expression $\sum_r (U_r(\overline{d}_r) - U_r'(\overline{d}_r)\overline{d}_r)$ is nonnegative, so we conclude that:

$$\frac{\sum_r U_r(\overline{d}_r)}{\sum_r U_r(d_r^S)} \geq \frac{\sum_r U_r'(\overline{d}_r)\overline{d}_r}{\left(\max_r U_r'(\overline{d}_r)\right) C},$$

since the right hand side of the expression above is less than or equal to 1. □

Let the vector $\mathbf{d}^G$ be the unique Nash equilibrium of the game with utility functions $U_1, \ldots, U_R$. We define a new collection of linear utility functions by:

$$\overline{U}_r(d_r) = U_r'(d_r^G) d_r.$$

Notice that the stationarity conditions (2.22)-(2.24) only involve the first derivatives of the utility functions U_r, $r = 1, \ldots, R$, at $\mathbf{d}^G$; thus, the unique Nash equilibrium of the game with utility functions $\overline{U}_1, \ldots, \overline{U}_R$ is given by $\mathbf{d}^G$ as well. Formally, $\mathbf{d}^G$ satisfies the stationarity conditions (2.22)-(2.24) for the family of utility functions $\overline{U}_1, \ldots, \overline{U}_R$. Furthermore, the optimal aggregate utility for this family of utility functions is given by $\left(\max_r U_r'(d_r^G)\right) C$. Applying Lemma 2.7 with $\overline{\mathbf{d}} = \mathbf{d}^G$, we thus see that the worst

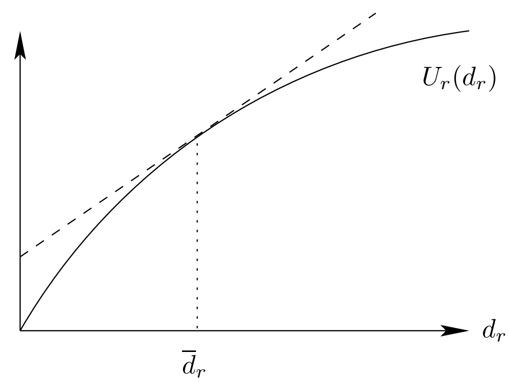

Figure 2-2. Proof of Lemma 2.7: We replace the utility function U_r by a linearization (dashed line) at the allocation $\overline{d}_r$. The linearization has the same value at $\overline{d}_r$, but is uniformly higher than U_r. Thus the optimal value of *SYSTEM* can only increase if we replace all utilities by their linearizations.

case efficiency loss occurs in the case of linear utility functions. We now proceed to calculate this efficiency loss.

Assume for the remainder of the proof, therefore, that U_r is linear, with $U_r(d_r) = \alpha_r d_r$, where $\alpha_r > 0$. Let $\mathbf{d}^G$ be the Nash equilibrium of the game with these utility functions. From the discussion in the preceding paragraph, the ratio of aggregate utility at the Nash equilibrium to aggregate utility at the optimal solution to *SYSTEM* is given by:

$$\frac{\sum_r \alpha_r d_r^G}{(\max_r \alpha_r)\, C}.$$

By scaling α_r and relabeling the users if necessary, we assume without loss of generality that $\max_r \alpha_r = \alpha_1 = 1$, and $C = 1$. To identify the worst case situation, we would like to find $\alpha_2, \ldots, \alpha_R$ such that $d_1^G + \sum_{r=2}^R \alpha_r d_r^G$, the total utility associated with the Nash equilibrium, is as small as possible; this results in the following optimization problem (with unknowns $d_1^G, \ldots, d_R^G, \alpha_2, \ldots, \alpha_R$):

$$\text{minimize} \quad d_1^G + \sum_{r=2}^R \alpha_r d_r^G \tag{2.26}$$

$$\text{subject to} \quad \alpha_r(1 - d_r^G) = 1 - d_1^G, \quad \text{if } d_r^G > 0; \tag{2.27}$$

$$\alpha_r \le 1 - d_1^G, \quad \text{if } d_r^G = 0; \tag{2.28}$$

$$\sum_r d_r^G = 1; \tag{2.29}$$

$$0 \le \alpha_r \le 1, \quad r = 2, \ldots, R; \tag{2.30}$$

$$d_r^G \ge 0, \quad r = 1, \ldots, R. \tag{2.31}$$

This optimization problem *chooses* linear utility functions with slopes less than or equal to 1 for players $2, \ldots, R$. The constraints in the problem require that given linear utility functions $U_r(d_r) = \alpha_r d_r$ for $r = 1, \ldots, R$, the allocation $\mathbf{d}^G$ must in fact be the unique Nash equilibrium allocation of the resulting game. As a result, the optimal objective function value is precisely the lowest possible aggregate utility achieved, among all such games. In addition, since $C = 1$, and the largest α_r is $\alpha_1 = 1$, the optimal aggregate utility is exactly 1; thus, the optimal objective function value of this problem also directly gives the worst case efficiency loss.

Suppose now $(\boldsymbol{\alpha}, \mathbf{d})$ is an optimal solution to (2.26)-(2.31) in which $n < R$ users, say users $r = R-n+1, \ldots, R$, have $d_r^G = 0$. Then the first $R-n$ coordinates of $\boldsymbol{\alpha}$ and $\mathbf{d}$ must be an optimal solution to the problem (2.26)-(2.31), with $R-n$ users. Therefore, in finding the worst case game, it suffices to assume that $d_r^G > 0$ for all $r = 2, \ldots, R$, and then consider the optimal objective function value for $R = 2, 3, \ldots$. This allows us to consider only the constraint:

$$\alpha_r(1 - d_r^G) = 1 - d_1^G. \tag{2.32}$$

This constraint then implies that $\alpha_r = (1 - d_1^G)/(1 - d_r^G)$. We will solve the resulting "reduced" optimization problem by decomposing it into two stages. First, we fix a choice of d_1^G and optimize over d_r^G, $r = 2, \ldots, R$; then, we choose the optimal value of d_1^G.

Given these observations, we fix d_1^G, and consider the following, simpler optimization problem:

$$\begin{aligned}
\text{minimize} \quad & d_1^G + \sum_{r=2}^{R} \frac{d_r^G(1 - d_1^G)}{1 - d_r^G} \\
\text{subject to} \quad & \sum_{r=2}^{R} d_r^G = 1 - d_1^G; \\
& 0 \leq d_r^G \leq d_1^G, \quad r = 2, \ldots, R.
\end{aligned}$$

Notice that we have substituted for α_r in the objective function. The constraint $\alpha_r \leq 1$ becomes equivalent to $d_r^G \leq d_1^G$ under the identification (2.32).

This simpler problem is only well defined if $d_1^G \geq 1/R$; otherwise the feasible region is empty—in other words, there exist no Nash equilibria with $d_1^G < 1/R$. If we assume that $d_1^G \geq 1/R$, then the feasible region is convex, compact, and nonempty, and the objective function is strictly convex in each of the variables d_r^G, $r = 2, \ldots, R$. This is sufficient to ensure that there exists a unique optimal solution as a function of

d_1^G; further, by symmetry, this optimal solution must be:

$$d_r^G = \frac{1 - d_1^G}{R - 1},$$

for $r = 2, \ldots, R$.

We now optimize over d_1^G. After substituting, we have the following optimization problem:

$$\begin{aligned} &\text{minimize} \quad && d_1^G + (1 - d_1^G)^2 \left(1 - \frac{1 - d_1^G}{R - 1}\right)^{-1} \\ &\text{subject to} \quad && \frac{1}{R} \le d_1^G \le 1. \end{aligned}$$

The objective function for the preceding optimization problem is decreasing in R for every value of d_1^G; in the limit where $R \to \infty$, the worst case efficiency loss is given by solving the following problem:

$$\begin{aligned} &\text{minimize} \quad && d_1^G + (1 - d_1^G)^2 \\ &\text{subject to} \quad && 0 \le d_1^G \le 1. \end{aligned}$$

It is simple to establish that the optimal solution to this problem occurs at $d_1^G = 1/2$, which yields a worst case aggregate utility of $3/4$, as required. This bound is tight in the limit where the number of users increases to infinity; using this fact, we obtain the tightness claimed in the theorem. □

The preceding theorem shows that in the worst case, aggregate utility falls by no more than 25% when users are able to anticipate the effects of their actions on the price of the link. Furthermore, this bound is essentially tight. In fact, it follows from the proof that the worst case consists of a link of capacity 1, where user 1 has utility $U_1(d_1) = d_1$, and all other users have utility $U_r(d_r) \approx d_r/2$ (when R is large). As $R \to \infty$, at the Nash equilibrium of this game user 1 receives a rate $d_1^G = 1/2$, while the remaining users uniformly split the rate $1 - d_1^G = 1/2$ among themselves, yielding an aggregate utility of $3/4$.

We note that a similar bound was observed by Roughgarden and Tardos for traffic routing games with affine link latency functions [108]. They found that the ratio of worst case Nash equilibrium cost to optimal cost was $4/3$. However, it remains an open question whether a relationship can be drawn between the two games; in particular, we note that while Theorem 2.6 holds even if the utility functions are nonlinear, Roughgarden and Tardos have shown that the efficiency loss due to selfish users in traffic routing may be arbitrarily high if link latency functions are nonlinear.

We conclude this section with a limit theorem, which shows that if the number of users grows to infinity while each user is allocated a negligible fraction of the resource, then the efficiency loss becomes negligible.

Corollary 2.8

Let $U_1, U_2, \ldots$ be a sequence of utility functions such that Assumption 2.1 is satisfied, and $\sup_r U_r'(0) < \infty$. Denote by $\mathbf{d}^G(R)$ and $\mathbf{d}^S(R)$ the unique optimal solution to GAME and any optimal solution to SYSTEM, respectively, when R users with utility functions $U_1, \ldots, U_R$ share a single link of capacity C. If:

$$\lim_{R\to\infty} d_r^G(R) = 0 \ \textit{for all } r,$$

then:

$$\lim_{R\to\infty} \frac{\sum_{r=1}^R U_r(d_r^G(R))}{\sum_{r=1}^R U_r(d_r^S(R))} = 1.$$

Proof. Let $\gamma = \sup_r U_r'(0) < \infty$. Let $\rho(R)$ denote the unique Nash equilibrium price when the utility functions are $U_1, \ldots, U_R$. We first show that $\sup_{R>1} \rho(R) \leq \gamma$. Suppose not; then choose R such that $\rho(R) > \gamma$. But then we must have $d_r^G(R) = 0$ for all $r = 1, \ldots, R$, which is impossible. Thus $\rho(R)$ remains bounded below γ; choose a convergent subsequence $\rho(R_k) \to \rho \leq \gamma$. We will show $\rho = \gamma$. To see this, note that for all R_k and all $r = 1, \ldots, R_k$, we have:

$$U_r'(d_r^G(R_k)) \left(1 - \frac{d_r^G(R_k)}{C}\right) \leq \rho(R_k).$$

Now as $R_k \to \infty$, the left hand side converges to $U_r'(0)$ for all r, and the right hand side converges to ρ. Thus $\rho \geq \sup_r U_r'(0) = \gamma$, so $\rho = \gamma$. Thus $\lim_{R\to\infty} \rho(R) = \gamma$.

By applying Lemma 2.7, we see that:

$$1 \geq \frac{\sum_{r=1}^R U_r(d_r^G(R))}{\sum_{r=1}^R U_r(d_r^S(R))} \geq \frac{\sum_{r=1}^R U_r'(d_r^G(R)) d_r^G(R)}{\left(\max_{r=1}^R U_r'(d_r^G(R))\right) C}.$$

Now if $d_r^G(R) > 0$, then we have $U_r'(d_r^G(R))(1 - d_r^G(R)/C) = \rho(R)$, so that $U_r'(d_r^G(R)) > \rho(R)$. Thus $\sum_{r=1}^R U_r'(d_r^G(R)) d_r^G(R) \geq \rho(R)C$. On the other hand, by concavity we have $U_r'(d_r^G(R)) \leq U_r'(0)$, so that $\max_{r=1}^R U_r'(d_r^G(R)) \leq \gamma$. We conclude that:

$$1 \geq \frac{\sum_{r=1}^R U_r(d_r^G(R))}{\sum_{r=1}^R U_r(d_r^S(R))} \geq \frac{\rho(R)}{\gamma}.$$

As $R \to \infty$, the right hand side converges to 1, which establishes the desired result. □

The previous result refers to a "competitive limit": as the system becomes large, if each user consumes an infinitesimal fraction of the resource, then the price anticipating behavior of the players does not adversely affect the system; as a result, it is *as if* all users were price takers, and the link begins to operate near an efficient allocation. Such results are frequently observed in oligopoly models; see, e.g., [82] for details.

2.3 Profit Maximizing Link Managers

Before continuing to our discussion of general network topologies, we will consider the effects of selfish link managers on the efficiency results above. We have assumed to this point that the link manager does not manipulate the pricing mechanism to his advantage. Specifically, we might expect the link manager to advertise a capacity $\hat{C}$ which is *strictly lower* than the true capacity C, with the aim of increasing the total revenues $\sum_r w_r$ received at the Nash equilibrium. In this section, we will explore a formal model to determine the consequences of such behavior.

We consider a two stage model. At the first stage, the link manager chooses an advertised capacity $\hat{C} > 0$, such that $\hat{C} \leq C$. At the second stage, R users take the advertised capacity $\hat{C}$ as given and compete for the link, where $R > 1$. Thus the R users have no knowledge of the true capacity C; the pricing mechanism of Section 2.1 is used to allocate the advertised capacity $\hat{C}$ among the users. We assume all utility functions are linear; let $U_r(d_r) = \alpha_r d_r$, where $\alpha_r > 0$. We let $\mathbf{w}(\hat{C})$ represent the unique Nash equilibrium at the second stage of the game (cf. Theorem 2.2), and let $\mathbf{d}(\hat{C})$ represent the associated rate allocation. The revenue to the link manager when the advertised capacity is $\hat{C}$ is given by $\sum_r w_r(\hat{C})$. We assume that the link manager knows exactly the dependence of the Nash equilibrium $\mathbf{w}(\hat{C})$ at the second stage on the choice of advertised capacity $\hat{C}$ made at the first stage. In this case, the link manager chooses $\hat{C}$ to maximize $\sum_r w_r(\hat{C})$ over $\hat{C} \in (0, C]$. We have the following theorem.

Theorem 2.9
Assume $R > 1$, and $U_r(d_r) = \alpha_r d_r$, where $\alpha_r > 0$. Then:

$$C = \arg \max_{0 < \hat{C} \leq C} \left[\sum_r w_r(\hat{C}) \right].$$

That is, advertising the true capacity C is the unique optimal strategy for the link manager.

Proof. The key to the proof is to show that the following equality holds for all $\hat{C} \in (0, C]$:

$$\sum_r w_r(\hat{C}) = \frac{\hat{C}}{C} \sum_r w_r(C).$$

This follows by examining the stationarity conditions (2.20)-(2.21), which imply that:

$$\alpha_r \left(1 - \frac{w_r(\hat{C})}{\sum_s w_s(\hat{C})}\right) = \frac{\sum_s w_s(\hat{C})}{\hat{C}}, \quad \text{if } w_r(\hat{C}) > 0; \tag{2.33}$$

$$\alpha_r \leq \frac{\sum_s w_s(\hat{C})}{\hat{C}}, \quad \text{if } w_r(\hat{C}) = 0. \tag{2.34}$$

Now define $w_r = C w_r(\hat{C})/\hat{C}$. Then the preceding relations imply:

$$\alpha_r \left(1 - \frac{w_r}{\sum_s w_s}\right) = \frac{\sum_s w_s}{C}, \quad \text{if } w_r > 0; \tag{2.35}$$

$$\alpha_r \leq \frac{\sum_s w_s}{C}, \quad \text{if } w_r = 0. \tag{2.36}$$

These are exactly the stationarity conditions (2.20)-(2.21), when the advertised capacity is C. Since the Nash equilibrium is unique, we must have $\mathbf{w} = \mathbf{w}(C)$. Thus:

$$\sum_r w_r(C) = \sum_r w_r = \frac{C}{\hat{C}} \sum_r w_r(\hat{C}).$$

Rewriting, we have:

$$\sum_r w_r(\hat{C}) = \frac{\hat{C}}{C} \sum_r w_r(C).$$

But now notice that $\sum_r w_r(\hat{C})$ is strictly increasing in $\hat{C}$, and maximized if and only if $\hat{C} = C$, as required. □

The preceding theorem shows that the optimal decision for a link manager is to always truthfully declare the capacity C, since this maximizes revenue. When combined with Theorem 2.6, we conclude that the efficiency loss will be no more than 25% when users are price anticipating, *and* the link manager is profit maximizing—provided that all utility functions are linear. Of course, the theorem is significantly dependent on the assumption that utility functions are linear. Indeed, the assumption of linear utility ensures that the revenue to the link manager is *scale invariant*—i.e., the revenue increases linearly in the advertised capacity $\hat{C}$. On the other hand, in general when utility is nonlinear, characterizing the dependence of the revenue on the advertised capacity is much more difficult, and thus characterizing optimal strategies for the link manager is not straightforward.

We may extend this scenario to a game where users have the choice of multiple parallel links which they can use. We label the links $1, \dots, J$, and let the true capacity of link j be C_j. Each link j is managed by a separate link manager, who chooses an advertised capacity $\hat{C}_j > 0$, such that $\hat{C}_j \leq C_j$. The strategy of user r is now a vector $\mathbf{w}_r = (w_{1r}, \ldots, w_{Jr})$, where w_{jr} represents the bid of user r to link j. When the advertised capacities are $\hat{\mathbf{C}} = (\hat{C}_1, \ldots, \hat{C}_J)$, and the composite strategy vector is $\mathbf{w}$, the resulting payoff to user r is:

$$Q_r(\mathbf{w}_r; \mathbf{w}_{-r}) = \alpha_r \left(\sum_{j: w_{jr} > 0} \frac{w_{jr}}{\sum_s w_{js}} \hat{C}_j \right) - \sum_j w_{jr}$$

$$= \sum_{j: w_{jr} > 0} \left(\alpha_r \frac{w_{jr}}{\sum_s w_{js}} \hat{C}_j - w_{jr} \right).$$

Since the utility is linear, the second equality above shows this game is *as if* J separate games are played by the R users, one at each link j. This implies there exists a unique Nash equilibrium $\mathbf{w}(\hat{\mathbf{C}})$. Furthermore, since the j games are independent from each other, the total revenue $\sum_r w_{jr}(\hat{\mathbf{C}})$ to link j depends only on $\hat{C}_j$. We thus have the following theorem; the proof is identical to the proof of Theorem 2.9.

Theorem 2.10
Assume $R > 1$, and $U_r(d_r) = \alpha_r d_r$, where $\alpha_r > 0$. Then for each j, independent of the value of $\hat{\mathbf{C}}_{-j}$:

$$C_j = \arg \max_{0 < \hat{C}_j \leq C} \left[\sum_r w_r(\hat{\mathbf{C}}) \right].$$

That is, advertising the true capacity C_j is the unique optimal strategy for the manager of link j, regardless of the strategies of the other link managers.

The previous theorem shows that as long as users' utilities are linear, advertising the true capacity will be optimal for all link managers in a network of parallel links. Note that in this case each of the links are *perfect substitutes* for each other. An interesting open question concerns determining the optimal strategies for profit maximizing link managers either when users' utility functions are nonlinear, or when the network topology is more complex. We also note that an interesting question concerns the implications for network performance when providers are profit maximizing, and are not constrained to use the pricing scheme of this chapter; one might ask what pricing scheme should be chosen by a network manager to maximize revenue, even if users are selfish. A recent paper of Ozdaglar and Acemoglu [97] investigates such a model when the link managers choose a price per unit rate sent through their link; under the assumption that users act as price takers, Ozdaglar and Acemoglu characterize the optimal pricing strategy for the link managers. These models suggest a problem of *optimal mechanism design*, as first proposed by Myerson [88]; a further investigation of this topic remains an interesting open question.

2.4 General Networks

In this section we will consider an extension of the single link model to general networks. We consider a network consisting of J links, numbered $1, \ldots, J$. Link j has a *capacity* given by $C_j > 0$; we let $\mathbf{C} = (C_1, \ldots, C_J)$ denote the vector of capacities. As before, a set of users numbered $1, \ldots, R$ shares this network of links. We assume there exists a set of paths through the network, numbered $1, \ldots, P$. By an abuse of notation, we will use J, R, and P to also denote the sets of links, users, and paths, respectively. Each path $p \in P$ uses a subset of the set of links J; if link j is used by path p, we will denote this by writing $j \in p$. Each user $r \in R$ has a collection of paths available through the network; if path p serves user r, we will denote this by writing $p \in r$. We will assume without loss of generality that paths are uniquely identified with users, so that for each path p there exists a unique user r such that $p \in r$. (There is no loss of generality because if two users share the same path, that is captured in our model by creating two paths which use exactly the same subset of links.) For notational convenience, we note that the links required by individual paths are captured by the *path-link incidence matrix* $\mathbf{A}$, defined by:

$$A_{jp} = \begin{cases} 1, & \text{if } j \in p; \\ 0, & \text{if } j \notin p. \end{cases}$$

Furthermore, we can capture the relationship between paths and users by the *path-user incidence matrix* $\mathbf{H}$, defined by:

$$H_{rp} = \begin{cases} 1, & \text{if } p \in r; \\ 0, & \text{if } p \notin r. \end{cases}$$

Note that by our assumption on paths, for each path p we have $H_{rp} = 1$ for exactly one user r.

Let $y_p \geq 0$ denote the rate allocated to path p, and let $d_r = \sum_{p \in r} y_p \geq 0$ denote the rate allocated to user r; using the matrix $\mathbf{H}$, we may write the relation between $\mathbf{d} = (d_r, r \in R)$ and $\mathbf{y} = (y_p, p \in P)$ as $\mathbf{Hy} = \mathbf{d}$. Any feasible rate allocation $\mathbf{y}$ must satisfy the following constraint:

$$\sum_{p: j \in p} y_p \leq C_j, \quad j \in J.$$

Using the matrix $\mathbf{A}$, we may write this constraint as $\mathbf{Ay} \leq \mathbf{C}$.

We continue to assume that user r receives a utility $U_r(d_r)$ from an amount of rate d_r, where the functions U_r satisfy Assumption 2.1. The natural generalization of the problem *SYSTEM* to a network context is given by the following optimization problem:

SYSTEM:

$$\text{maximize} \quad \sum_r U_r(d_r) \tag{2.37}$$

$$\text{subject to} \quad \mathbf{A}\mathbf{y} \le \mathbf{C}; \tag{2.38}$$

$$\mathbf{H}\mathbf{y} = \mathbf{d}; \tag{2.39}$$

$$y_p \ge 0, \ \ p \in P. \tag{2.40}$$

Since the objective function is continuous and the feasible region is compact, an optimal solution $\mathbf{y}$ exists; since the feasible region is also convex, if the functions U_r are strictly concave, then the optimal vector $\mathbf{d} = \mathbf{H}\mathbf{y}$ is uniquely defined (though $\mathbf{y}$ need not be unique). As in Section 2.2, we will use the optimal solution to *SYSTEM* as a benchmark for the outcome of the network game.

We now define the resource allocation mechanism for this network setting. The natural extension of the single link model is defined as follows. Each user r submits a *bid* w_{jr} for each link j; this defines a strategy for user r given by $\mathbf{w}_r = (w_{jr}, j \in J)$, and a composite strategy vector given by $\mathbf{w} = (\mathbf{w}_1, \ldots, \mathbf{w}_R)$. We then assume that each link takes these bids as input, and uses the pricing scheme developed in Section 2.1. Formally, each link sets a price $\mu_j(\mathbf{w})$, given by:

$$\mu_j(\mathbf{w}) = \frac{\sum_r w_{jr}}{C_j}. \tag{2.41}$$

As before, we assume the rate allocated to a user is proportional to his payment. We denote by $x_{jr}(\mathbf{w})$ the rate allocated to user r by link j; we thus have:

$$x_{jr}(\mathbf{w}) = \begin{cases} \dfrac{w_{jr}}{\mu_j(\mathbf{w})}, & \text{if } w_{jr} > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{2.42}$$

We define the vector $\mathbf{x}_r(\mathbf{w})$ by:

$$\mathbf{x}_r(\mathbf{w}) = (x_{jr}(\mathbf{w}), j \in J).$$

Now given any vector $\overline{\mathbf{x}}_r = (\overline{x}_{jr}, j \in J)$, we define $d_r(\overline{\mathbf{x}}_r)$ to be the optimal value of the following optimization problem:

$$\text{maximize} \quad \sum_{p \in r} y_p \tag{2.43}$$

$$\text{subject to} \quad \sum_{p \in r : j \in p} y_p \le \overline{x}_{jr}, \;\; j \in J; \tag{2.44}$$

$$y_p \ge 0, \;\; p \in r. \tag{2.45}$$

Given the strategy vector $\mathbf{w}$, we then define the rate allocated to user r as $d_r(\mathbf{x}_r(\mathbf{w}))$. To gain some intuition for this allocation mechanism, notice that when the vector of bids is $\mathbf{w}$, user r is allocated a rate $x_{jr}(\mathbf{w})$ at each link j. Since the utility to user r is nondecreasing in the total amount of rate allocated, user r's utility is maximized if he solves the preceding optimization problem, which is a *max-flow* problem constrained by the rate x_{jr} available at each link j. In other words, user r is allocated the maximum possible rate $d_r(\mathbf{x}_r(\mathbf{w}))$, given that each link j has granted him rate $x_{jr}(\mathbf{w})$. (Note that this is not the same as the mechanism proposed by Kelly in [62], where users submit only a single *total payment* to the network; we explore the consequences of this difference further in Section 2.4.3.)

Define the notation $\mathbf{w}_{-r} = (\mathbf{w}_1, \ldots, \mathbf{w}_{r-1}, \mathbf{w}_{r+1}, \ldots, \mathbf{w}_R)$. Based on the definition of $d_r(\mathbf{x}_r(\mathbf{w}))$ above, the payoff to user r is given by:

$$Q_r(\mathbf{w}_r; \mathbf{w}_{-r}) = U_r\big(d_r(\mathbf{x}_r(\mathbf{w}))\big) - \sum_j w_{jr}. \tag{2.46}$$

A *Nash equilibrium* of the game defined by $(Q_1, \ldots, Q_R)$ is a vector $\mathbf{w} \ge 0$ such that for all r:

$$Q_r(\mathbf{w}_r; \mathbf{w}_{-r}) \ge Q_r(\overline{\mathbf{w}}_r; \mathbf{w}_{-r}), \;\; \text{for all } \overline{\mathbf{w}}_r \ge 0. \tag{2.47}$$

Although this pricing scheme is very natural, the fact that the payoff Q_r may be discontinuous can prevent existence of a Nash equilibrium, as we first observed in Example 2.1. Although we were able to prove a Nash equilibrium exists with $R > 1$ users for the single link case, the following example shows that Nash equilibria may not exist in the network context even if $R > 1$.

Example 2.2

Consider an example consisting of two links, labeled $j = 1$, and $j = 2$. The first link has capacity C_1, and the second link has capacity $C_2 > C_1$, as depicted in Figure 2-3. The system consists of R users, where we assume that each user r has a strictly increasing, concave, continuous utility function U_r. For this example, we will assume each user r is identified with a single path consisting of both links 1 and 2. This simplifies the

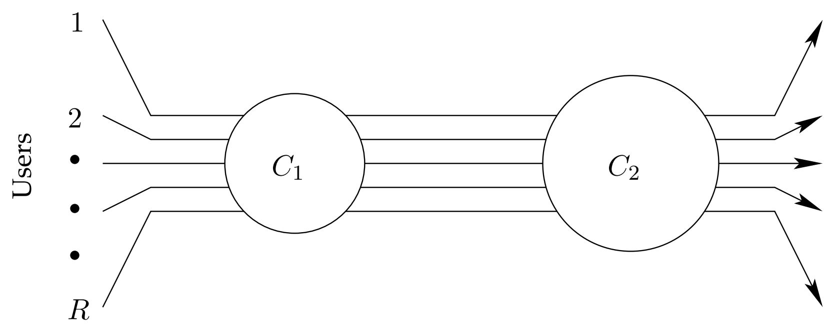

Figure 2-3. Example 2.2: Link 1 has capacity C_1, and link 2 has capacity C_2, where $C_1 < C_2$. Each one of R users requires service from both links.

analysis, since the optimal solution to the problem (2.43)-(2.45) is then given by:

$$d_r(\mathbf{x}_r(\mathbf{w})) = \min\{x_{1r}(\mathbf{w}), x_{2r}(\mathbf{w})\}.$$

We will show that no Nash equilibrium exists for this network. Suppose, to the contrary, that $\mathbf{w}$ is a Nash equilibrium. We first show that $\sum_r w_{jr} > 0$, for $j = 1, 2$. If not, then all users are allocated zero rate. First suppose that $\sum_r w_{jr} = 0$ for both $j = 1, 2$. Then any user r can profitably deviate by infinitesimally increasing w_{1r} and w_{2r}, say by $\Delta > 0$; this deviation will give user r rate $\min\{C_1, C_2\} = C_1$, and increase the total payment by 2Δ. For small enough Δ, this will strictly improve the payoff of player r; thus no Nash equilibrium exists where $\sum_r w_{jr} = 0$ for both $j = 1, 2$. A similar argument follows if $\sum_r w_{1r} = 0$, but $\sum_r w_{2r} > 0$: in this case, for any user r such that $w_{2r} > 0$, a profitable deviation exists where w_{2r} is reduced to zero; this leaves user r's rate allocation unchanged at zero, while reducing his total payment to the network. Symmetrically, the same argument may be used when $\sum_r w_{1r} > 0$, and $\sum_r w_{2r} = 0$. As a result, we conclude that if $\mathbf{w}$ is a Nash equilibrium, we must have $\sum_r w_{jr} > 0$ for both $j = 1, 2$.

Now note that (trivially) we have the relations:

$$\sum_r \frac{w_{1r}}{\sum_s w_{1s}} C_1 = C_1; \text{ and } \sum_r \frac{w_{2r}}{\sum_s w_{2s}} C_2 = C_2.$$

Since $C_1 < C_2$, there must exist at least one user r for whom $(w_{1r}C_1)/(\sum_s w_{1s}) < (w_{2r}C_2)/(\sum_s w_{2s})$. Recall that user r is allocated a total rate equal to:

$$\min\left\{\frac{w_{1r}}{\sum_s w_{1s}} C_1, \frac{w_{2r}}{\sum_s w_{2s}} C_2\right\}.$$

As a result, user r can profitably deviate by reducing w_{2r}, since this reduces his payment, without altering his rate allocation. Thus no such vector $\mathbf{w}$ can be a Nash equilibrium. □

As will be seen in the following development, the issue in the previous example is that link 2 is not a bottleneck in the network (since $C_1 < C_2$, link 2 will never be fully utilized). As a result, as long as the total payment $\sum_s w_{2s}$ to link 2 is strictly positive, there will always be some user r who is overpaying—i.e., this user could profitably deviate by reducing w_{2r}. Thus the only equilibrium outcome is one where the total payment to link 2 becomes zero; but in this case, because of the discontinuity in the payoff function defined in (2.46) (or, more precisely, the discontinuity in (2.42)), all users are allocated zero rate. In fact, by a similar argument it is possible to see that a competitive equilibrium need not exist in general either. Considering the same model as Example 2.2, one can show that at any competitive equilibrium the price μ_2 at link 2 must be zero; however, in that case the payoff to any user r is not well defined for $w_r > 0$.

We will see in the following section that a resolution to this problem can be found if users are allowed to request and be allocated a nonzero rate from links for which the total payment is zero. We show that Nash equilibria are always guaranteed to exist for this "extended" game; furthermore, we show that any Nash equilibrium for the game defined by $(Q_1, \ldots, Q_R)$ corresponds in a natural way to a Nash equilibrium of the extended game. In Section 2.4.2, we show that the aggregate utility at any Nash equilibrium of the extended game is no less than 3/4 times the *SYSTEM* optimal aggregate utility, matching the result achieved for the single link game. Finally, in Section 2.4.3 we consider the relationship between the network pricing proposal given here and the original pricing proposal of Kelly [62].

■ 2.4.1 An Extended Game

In this section, we consider an extended game, where users not only submit bids, but also rate requests. We consider an allocation mechanism under which the rate requests are only taken into account by a link when the total payment to that link is zero. This behavior ensures that when a link is not congested (as in Example 2.2), or is not in sufficient demand (as in Example 2.1), users may still be allocated a nonzero rate on that link. In particular, this modification addresses the degeneracies which arise due to the discontinuity of Q_r in the original definition of the network game. We will show that Nash equilibria always exist for this extended game. (We note that extended strategy spaces have also proven fruitful for other games with payoff discontinuities; see, e.g., [57].)

Formally, we suppose that the strategy of user r includes a *rate request* $\phi_{jr} \geq 0$ at each link j; that is, the strategy of user r is a vector $\boldsymbol{\sigma}_r = (\boldsymbol{\phi}_r, \mathbf{w}_r)$, where $\boldsymbol{\phi}_r = (\phi_{jr}, j \in J)$, and $\mathbf{w}_r = (w_{jr}, j \in J)$, as before. We will write $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_R)$

to denote the composite strategy vector of all players; and we will use the notation $\boldsymbol{\sigma}_{-r} = (\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_{r-1}, \boldsymbol{\sigma}_{r+1}, \ldots, \boldsymbol{\sigma}_R)$ to denote all components of $\boldsymbol{\sigma}$ other than $\boldsymbol{\sigma}_r$. We now suppose that each link j provides a rate $x_{jr}(\boldsymbol{\sigma})$ to user r, which is determined as follows:

1. If $\sum_s w_{js} > 0$, then:
$$x_{jr}(\boldsymbol{\sigma}) = \frac{w_{jr}}{\sum_s w_{js}} C_j. \tag{2.48}$$

2. If $\sum_s w_{js} = 0$, and $\sum_s \phi_{js} \leq C_j$, then:
$$x_{jr}(\boldsymbol{\sigma}) = \phi_{jr}. \tag{2.49}$$

3. If $\sum_s w_{js} = 0$, and $\sum_s \phi_{js} > C_j$, then:
$$x_{jr}(\boldsymbol{\sigma}) = 0. \tag{2.50}$$

In the first instance, when link j receives a positive payment from the users, rate is allocated in proportion to the bids. The second two cases apply only when the total payment to link j is zero; in this event, if the total requested rate is less than the capacity C_j, then the requests are granted. However, if the total requested rate exceeds capacity, no rate is allocated. We note here that the precise definition in case 3 above is not essential; any mechanism which splits the capacity C_j according to a preset deterministic rule will lead to the same results below. For example, if requests exceed capacity, a link may choose to allocate the same rate to all users who share the link; or the link may choose to allocate all the entire capacity to some predetermined "preferred" user.

It is straightforward to check, using methods similar to the proof of Theorem 2.1, that a competitive equilibrium exists under this pricing mechanism, and any competitive equilibrium is an optimal solution to *SYSTEM*. A competitive equilibrium consists of a strategy vector $\boldsymbol{\sigma}_r$ for each user r, as well as a price μ_j for each link j, such that $\mu_j = \sum_s w_{js}/C_j$ for all j, and each user r has optimally chosen his strategy $\boldsymbol{\sigma}_r$ while taking the link prices μ_j as given.

As before, we define:
$$\mathbf{x}_r(\boldsymbol{\sigma}) = (x_{jr}(\boldsymbol{\sigma}), j \in J).$$

The rate of user r is then $d_r(\mathbf{x}_r(\boldsymbol{\sigma}))$ (where d_r is defined as the optimal value to the optimization problem (2.43)-(2.45)). The payoff T_r to user r is given by:
$$T_r(\boldsymbol{\sigma}_r; \boldsymbol{\sigma}_{-r}) = U_r\big(d_r(\mathbf{x}_r(\boldsymbol{\sigma}))\big) - \sum_j w_{jr}. \tag{2.51}$$

(Note that this is an abuse of notation in the definition of $\mathbf{x}_r$ and x_{jr}, since we previously had defined them as functions of $\mathbf{w}$. However, the definition in use will be clear from the argument of the function.)

A *Nash equilibrium* of the game defined by $(T_1, \dots, T_R)$ is a vector $\boldsymbol{\sigma} \geq 0$ such that for all r:

$$T_r(\boldsymbol{\sigma}_r; \boldsymbol{\sigma}_{-r}) \geq T_r(\overline{\boldsymbol{\sigma}}_r; \boldsymbol{\sigma}_{-r}), \text{ for all } \overline{\boldsymbol{\sigma}}_r \geq 0. \tag{2.52}$$

Because we have extended the strategy space, in fact it is possible to prove existence of Nash equilibria under weaker conditions than those required in Section 2.1.2. In particular, we no longer require U_r to be strictly increasing or differentiable, as in the previous development. These modifications are summarized in the following assumption.

Assumption 2.2
For each r, the utility function $U_r(d_r)$ is concave, nondecreasing, and continuous over the domain $d_r \geq 0$.

From (2.48)-(2.50), the rate requests are only considered if the total payment to a link is zero. This fact leads us to expect that any Nash equilibrium of the original game defined by $(Q_1, \dots, Q_R)$ is also a Nash equilibrium of the new game defined by $(T_1, \dots, T_R)$, since in this case the rate requests should be meaningless. This is indeed true, as we show in the following theorem.

Theorem 2.11
Suppose that Assumption 2.2 is satisfied, and that $\mathbf{w}$ is a strategy vector for the game defined by $(Q_1, \dots, Q_R)$. For each user r, define:

$$\phi_{jr} = x_{jr}(\mathbf{w}) = \begin{cases} \dfrac{w_{jr}}{\sum_s w_{js}} C_j, & \textit{if } w_{jr} > 0; \\ 0, & \textit{otherwise}. \end{cases}$$

For each user r, let $\boldsymbol{\sigma}_r = (\boldsymbol{\phi}_r, \mathbf{w}_r)$. Then user r receives the same payoff in either game:

$$T_r(\boldsymbol{\sigma}_r; \boldsymbol{\sigma}_{-r}) = Q_r(\mathbf{w}_r; \mathbf{w}_{-r}).$$

Furthermore, if $\mathbf{w}$ is a Nash equilibrium of the game defined by $(Q_1, \dots, Q_R)$, then $\boldsymbol{\sigma}$ is a Nash equilibrium of the game defined by $(T_1, \dots, T_R)$.

Proof. We will refer to the game defined by $(Q_1, \dots, Q_R)$ as the "original game," and the game defined by $(T_1, \dots, T_R)$ as the "extended game." We first note that given the definition of ϕ_{jr} above, we have the identity $x_{jr}(\boldsymbol{\sigma}) = x_{jr}(\mathbf{w})$ for each link j; that is, the allocation from link j to user r in the extended game is identical to the allocation made by link j in the original game. Furthermore, the total payment made by user r remains unchanged in the extended game. Thus the payoff to user r is the same in both games, under the mapping from $\mathbf{w}$ to $\boldsymbol{\sigma}$ defined in the statement of the theorem.

Now suppose that $\mathbf{w}$ is a Nash equilibrium of the original game, and define $\boldsymbol{\sigma}$ as in the statement of the theorem. For each link j and each user r, define $W_{jr} = \sum_{s \neq r} w_{js}$. Suppose there exists a strategy vector $\overline{\boldsymbol{\sigma}}_r = (\overline{\boldsymbol{\phi}}_r, \overline{w}_r)$ such that:

$$U_r\big(d_r(\mathbf{x}_r(\overline{\boldsymbol{\sigma}}_r, \boldsymbol{\sigma}_{-r}))\big) - \sum_j \overline{w}_{jr} > U_r\big(d_r(\mathbf{x}_r(\boldsymbol{\sigma}))\big) - \sum_j w_{jr}.$$

Fix $\varepsilon > 0$. For each j, we define $\hat{w}_{jr} = \overline{w}_{jr}$ if $W_{jr} > 0$, and $\hat{w}_{jr} = \varepsilon$ if $W_{jr} = 0$. Then:

$$x_{jr}(\hat{\mathbf{w}}_r, \mathbf{w}_{-r}) \geq x_{jr}(\overline{\boldsymbol{\sigma}}_r, \boldsymbol{\sigma}_{-r}).$$

The preceding inequality follows because from each link $j \in r$ with $W_{jr} = 0$, user r is allocated the entire capacity C_j in return for the payment of $\varepsilon > 0$. From this we may conclude that:

$$d_r(\mathbf{x}_r(\hat{\mathbf{w}}_r, \mathbf{w}_{-r})) \geq d_r(\mathbf{x}_r(\overline{\boldsymbol{\sigma}}_r, \boldsymbol{\sigma}_{-r})).$$

Now as $\varepsilon \to 0$, we have $\lim_{\varepsilon \to 0} \sum_j \hat{w}_{jr} \leq \sum_j \overline{w}_{jr}$. Thus for sufficiently small $\varepsilon > 0$, we will have:

$$\begin{aligned}
U_r\big(d_r(\mathbf{x}_r(\hat{\mathbf{w}}_r, \mathbf{w}_{-r}))\big) - \sum_j \hat{w}_{jr} &\geq U_r\big(d_r(\mathbf{x}_r(\overline{\boldsymbol{\sigma}}_r, \boldsymbol{\sigma}_{-r}))\big) - \sum_j \hat{w}_{jr} \\
&> U_r\big(d_r(\mathbf{x}_r(\boldsymbol{\sigma}))\big) - \sum_j w_{jr} \\
&= U_r\big(d_r(\mathbf{x}_r(\mathbf{w}))\big) - \sum_j w_{jr}.
\end{aligned}$$

Thus the vector $\hat{\mathbf{w}}_r = (\hat{w}_{jr}, j \in r)$ is a profitable deviation for user r in the original game, a contradiction. Therefore no profitable deviation exists for user r in the extended game. We conclude $\boldsymbol{\sigma}$ is a Nash equilibrium for the extended game, as required. □

The preceding theorem shows that any Nash equilibrium of the original game corresponds naturally to a Nash equilibrium of the extended game. To construct a partial converse to this result, suppose that we are given a Nash equilibrium $\boldsymbol{\sigma} = (\boldsymbol{\phi}, \mathbf{w})$ of the extended game, but that $\sum_r w_{jr} > 0$ for all links j. We first note that for each link j, at least two distinct users submit positive bids. If not, then there is some link j where a single user r submits a positive bid—but this user can leave his rate allocation unchanged and reduce his payment by lowering the bid submitted to link j. Thus we conclude that for each link j and each user r, the payment by all other users $\sum_{s \neq r} w_{js}$ is positive. This ensures the rate requests $\boldsymbol{\phi}_r$ do not have any effect on the rate allocation made to user r, so that the payoffs are determined entirely by the bid vectors $\mathbf{w}_r$, for $r \in R$. This is sufficient to conclude that $\mathbf{w}$ must actually be a Nash equilibrium for the original game. To summarize, we have shown that whenever all link prices are positive at a Nash equilibrium in the extended game, then in fact we have a Nash equilibrium for the original game as well.

We now turn our attention to showing that a Nash equilibrium always exists for the extended game.

Theorem 2.12
Suppose that Assumption 2.2 is satisfied. Then a Nash equilibrium exists for the game defined by $(T_1, \dots, T_R)$.

Proof. Our technique is to consider a perturbed version of the *original* game, where a "virtual" user submits a bid of $\varepsilon > 0$ to each link j in the network. Formally, this means that at a bid vector $\mathbf{w}$, user r is allocated a rate $x^{\varepsilon}_{jr}(\mathbf{w})$ at link j, given by:

$$x^{\varepsilon}_{jr}(\mathbf{w}) = \frac{w_{jr}}{\varepsilon + \sum_s w_{js}} C_j.$$

We define the vector $\mathbf{x}^{\varepsilon}_r(\mathbf{w}) = (x^{\varepsilon}_{jr}(\mathbf{w}), j \in J)$, and the rate attained by user r is then $d_r(\mathbf{x}^{\varepsilon}_r(\mathbf{w}))$, where d_r is the optimal value to the optimization problem (2.43)-(2.45).

The modified allocation defined by $\mathbf{x}^{\varepsilon}_r$ was also considered by Maheswaran and Basar in the context of a single link [79]; we will use this allocation mechanism to prove existence for our game by taking a limit as $\varepsilon \to 0$. Our approach will be to first apply standard fixed point techniques to establish existence of a Nash equilibrium $\mathbf{w}^{\varepsilon}$ for this perturbed game, with an associated allocation to each user given by $\mathbf{x}^{\varepsilon}_r(\mathbf{w}^{\varepsilon})$. We will then show that $\mathbf{w}^{\varepsilon}$ and $\mathbf{x}^{\varepsilon}_r(\mathbf{w}^{\varepsilon})$ (for each r) lie in compact sets, respectively. If we then choose $\mathbf{w}$ and $\boldsymbol{\phi} = (\boldsymbol{\phi}_r, r \in R)$ as limit points when $\varepsilon \to 0$, we will find that $(\mathbf{w}, \boldsymbol{\phi})$ is a Nash equilibrium of the extended game.

Step 1: A Nash equilibrium $\mathbf{w}^{\varepsilon}$ *exists in the perturbed game.* We first observe that since $\varepsilon > 0$, $x^{\varepsilon}_{jr}(\mathbf{w})$ is a continuous, strictly concave, and strictly increasing function of $w_{jr} \geq 0$ (in particular, there is no longer any discontinuity in the rate allocation at $w_{jr} = 0$). Furthermore, since d_r is defined as the maximal objective value of a linear program, $d_r(\mathbf{x}_r)$ is concave and continuous as a function of $\mathbf{x}_r$ ([15], Section 5.2); and if $x_{jr} \geq \overline{x}_{jr}$ for all j, then clearly $d_r(\mathbf{x}_r) \geq d_r(\overline{\mathbf{x}}_r)$, i.e., d_r is *nondecreasing* (this follows immediately from the problem (2.43)-(2.45)).

We will now combine these facts to show that user r's payoff in this perturbed game is concave as a function of $\mathbf{w}_r$, and continuous as a function of the composite strategy $\mathbf{w}$. The payoff in the perturbed game, denoted Q^{ε}_r, is given by:

$$Q^{\varepsilon}_r(\mathbf{w}_r; \mathbf{w}_{-r}) = U_r\big(d_r(\mathbf{x}^{\varepsilon}_r(\mathbf{w}))\big) - \sum_{j \in r} w_{jr}.$$

Continuity of Q^{ε}_r as a function of $\mathbf{w}$ follows immediately from continuity of x^{ε}_{jr}, d_r, and U_r. To show that Q_r^ε is concave as a function of $\mathbf{w}_r$, it suffices to show that $U_r(d_r(\mathbf{x}_r^\varepsilon(\mathbf{w}_r, \mathbf{w}_{-r})))$ is a concave function of $\mathbf{w}_r$. Since for each j the function x_{jr}^ε is concave in w_{jr}, and does not depend on w_{kr} for $k \neq j$, we conclude that each component of $\mathbf{x}_r^\varepsilon(\mathbf{w}_r, \mathbf{w}_{-r})$ is a concave function of $\mathbf{w}_r$. If we fix the bids of the other players as $\mathbf{w}_{-r}$, then since d_r is nondecreasing and concave in its argument, we have for any two bid vectors $\mathbf{w}_r, \overline{\mathbf{w}}_r$, and δ such that $0 \leq \delta \leq 1$:

$$\begin{aligned} d_r(\mathbf{x}_r^\varepsilon(\delta \mathbf{w}_r + (1-\delta)\overline{\mathbf{w}}_r, \mathbf{w}_{-r})) &\geq d_r(\delta \mathbf{x}_r^\varepsilon(\mathbf{w}_r, \mathbf{w}_{-r}) + (1-\delta)\mathbf{x}_r^\varepsilon(\overline{\mathbf{w}}_r, \mathbf{w}_{-r})) \\ &\geq \delta d_r(\mathbf{x}_r^\varepsilon(\mathbf{w}_r, \mathbf{w}_{-r})) + (1-\delta) d_r(\mathbf{x}_r^\varepsilon(\overline{\mathbf{w}}_r, \mathbf{w}_{-r})). \end{aligned}$$

We now apply the fact that U_r is nondecreasing and concave to conclude that:

$$\begin{aligned} U_r\big(d_r(\mathbf{x}_r^\varepsilon(\delta \mathbf{w}_r + (1-\delta)\overline{\mathbf{w}}_r, \mathbf{w}_{-r}))\big) &\geq U_r\big(\delta d_r(\mathbf{x}_r^\varepsilon(\mathbf{w}_r, \mathbf{w}_{-r})) + (1-\delta) d_r(\mathbf{x}_r^\varepsilon(\overline{\mathbf{w}}_r, \mathbf{w}_{-r}))\big) \\ &\geq \delta U_r\big(d_r(\mathbf{x}_r^\varepsilon(\mathbf{w}_r, \mathbf{w}_{-r}))\big) + \\ &\qquad (1-\delta) U_r\big(d_r(\mathbf{x}_r^\varepsilon(\overline{\mathbf{w}}_r, \mathbf{w}_{-r}))\big). \end{aligned}$$

Thus user r's payoff function $Q_r^\varepsilon(\mathbf{w}_r; \mathbf{w}_{-r})$ is concave in $\mathbf{w}_r$.

Finally, we observe that in searching for a Nash equilibrium of the perturbed game defined by $(Q_1^\varepsilon, \ldots, Q_R^\varepsilon)$, we can restrict the strategy space of each user to a compact, convex subset of $\mathbb{R}^J$. To see this, fix a user r, and choose $B_r > U_r(\sum_j C_j) - U_r(0)$. When user r sets $\mathbf{w}_r = 0$, his payoff is $U_r(0)$. On the other hand, the maximum rate user r can be allocated from the network is bounded above by $\sum_j C_j$; and thus, if user r chooses any strategy $\mathbf{w}_r$ such that $\sum_j w_{jr} > B_r$, then regardless of the strategies $\mathbf{w}_{-r}$ of all other players, we have:

$$U_r(d_r(\mathbf{x}_r^\varepsilon(\mathbf{w}_r, \mathbf{w}_{-r})) - \sum_j w_{jr} \leq U_r(\sum_j C_j) - B_r < U_r(0).$$

Thus, if we define the compact set $S_r = \{\mathbf{w}_r : \sum_j w_{jr} \leq B_r\}$, we observe that user r would never choose a strategy vector that lies outside S_r; this allows us to restrict the strategy space of user r to the set S_r.

The game defined by $(Q_1^\varepsilon, \ldots, Q_R^\varepsilon)$ together with the strategy spaces $(S_1, \ldots, S_R)$ is then a *concave R-person game*: each payoff function is continuous in the composite strategy vector $\mathbf{w}$; Q_r^ε is concave in $\mathbf{w}_r$; and the strategy space of each user r is a compact, convex, nonempty subset of $\mathbb{R}^J$. Applying Rosen's existence theorem [104] (proven using Kakutani's fixed point theorem), we conclude that a Nash equilibrium $\mathbf{w}^\varepsilon$ exists for this game.

Step 2: There exists a limit point $\boldsymbol{\sigma} = (\boldsymbol{\phi}, \mathbf{w})$ of the Nash equilibria of the perturbed games. For each user r, define $\phi_{jr}^\varepsilon = x_{jr}^\varepsilon(\mathbf{w}^\varepsilon)$. Let $\boldsymbol{\phi}_r^\varepsilon = (\phi_{jr}^\varepsilon, j \in J)$, and $\boldsymbol{\phi}^\varepsilon = (\boldsymbol{\phi}_r^\varepsilon, r \in R)$. We now note that for all $\varepsilon > 0$, the pair $(\boldsymbol{\phi}^\varepsilon, \mathbf{w}^\varepsilon)$ lies in a compact subset of Euclidean space.

To see this, note that $\mathbf{w}^\varepsilon$ lies in the compact set $S_1 \times \cdots \times S_R$, and that $0 \le \phi_{jr}^\varepsilon \le C_j$ for all j and r. Thus, there exists a sequence $\varepsilon_k \to 0$ such that the sequence $(\boldsymbol{\phi}^{\varepsilon_k}, \mathbf{w}^{\varepsilon_k})$ converges to some $\boldsymbol{\sigma} = (\boldsymbol{\phi}, \mathbf{w})$, where $\mathbf{w} \in S_1 \times \cdots \times S_R$ and $0 \le \phi_{jr} \le C_j$.

We expect that at the limit point $\boldsymbol{\sigma}$, the rates allocated to each user are the limits of the rates allocated in the perturbed games. Formally, we show that we have:

$$x_{jr}(\boldsymbol{\sigma}) = \lim_{k\to\infty} x_{jr}^{\varepsilon_k}(\mathbf{w}^{\varepsilon_k}). \tag{2.53}$$

Fix a link j, and suppose that $\sum_r w_{jr} = 0$. By definition, $\phi_{jr} = \lim_{k\to\infty} x_{jr}^{\varepsilon_k}(\mathbf{w}^{\varepsilon_k})$ for each r. We thus only need to show that $x_{jr}(\boldsymbol{\sigma}) = \phi_{jr}$ for each r, which follows from the rate allocation mechanism since:

$$\sum_r \phi_{jr} = \lim_{k\to\infty} \sum_r x_{jr}^{\varepsilon_k}(\mathbf{w}^{\varepsilon_k}) \le C_j.$$

On the other hand, suppose that $\sum_r w_{jr} > 0$. In this case, we have that $x_{jr}(\boldsymbol{\sigma}) = (w_{jr}C_j)/(\sum_s w_{js}) = \lim_{k\to\infty} x_{jr}^{\varepsilon_k}(\mathbf{w}^{\varepsilon_k})$ for each r, as required.

Step 3: The vector $\boldsymbol{\sigma}$ is a Nash equilibrium of the extended game. Suppose $\boldsymbol{\sigma}$ is not a Nash equilibrium of the extended game; then there exists some user r, and a strategy vector $\overline{\boldsymbol{\sigma}}_r = (\overline{\boldsymbol{\phi}}_r, \overline{\mathbf{w}}_r)$, such that $T_r(\overline{\boldsymbol{\sigma}}_r; \boldsymbol{\sigma}_{-r}) > T_r(\boldsymbol{\sigma})$. Our goal will be to show that in this case, for sufficiently small $\varepsilon > 0$, a profitable deviation exists for user r from the strategy vector $\mathbf{w}_r^\varepsilon$ (i.e., from the chosen Nash equilibrium for the game defined by $Q_1^\varepsilon, \ldots, Q_R^\varepsilon$).

For fixed $\varepsilon > 0$, we now construct a new strategy vector $\overline{\mathbf{w}}_r^\varepsilon$ for user r. First fix a link j such that $W_{jr} > 0$; we then define $\overline{w}_{jr}^\varepsilon > 0$ by:

$$\overline{w}_{jr}^\varepsilon = \frac{W_{jr}^\varepsilon + \varepsilon}{W_{jr}} \overline{w}_{jr}.$$

Observe that with this definition, as $k \to \infty$, we have $\overline{w}_{jr}^{\varepsilon_k} \to \overline{w}_{jr}$. We also have:

$$\frac{\overline{w}_{jr}}{\overline{w}_{jr} + W_{jr}} C_j = \frac{\overline{w}_{jr}^\varepsilon}{\overline{w}_{jr}^\varepsilon + W_{jr}^\varepsilon + \varepsilon} C_j.$$

This implies that $x_{jr}(\overline{\boldsymbol{\sigma}}_r, \boldsymbol{\sigma}_{-r}) = x_{jr}^\varepsilon(\overline{\mathbf{w}}_r^\varepsilon, \mathbf{w}_{-r}^\varepsilon)$, regardless of how we define the remaining components of the vector $\overline{\mathbf{w}}_r^\varepsilon$.

To complete this definition, suppose now that we fix a link j such that $W_{jr} = 0$. In this case we define $\overline{w}_{jr}^\varepsilon = \sqrt{W_{jr}^\varepsilon + \varepsilon}$. (The specific form is not important here; for the proof we only require that when $W_{jr} = 0$, we have $\overline{w}_{jr}^\varepsilon / (W_{jr}^\varepsilon + \varepsilon) \to \infty$ as $\varepsilon \to 0$.)

Then we have $\overline{w}_{jr}^{\varepsilon_k} \to 0$ as $k \to \infty$. Furthermore:

$$x_{jr}^{\varepsilon}(\overline{\mathbf{w}}_r^{\varepsilon}, \mathbf{w}_{-r}^{\varepsilon}) = \frac{\sqrt{W_{jr}^{\varepsilon} + \varepsilon}}{\sqrt{W_{jr}^{\varepsilon} + \varepsilon} + W_{jr}^{\varepsilon} + \varepsilon} C_j.$$

Since $W_{jr}^{\varepsilon_k} + \varepsilon_k \to 0$ as $k \to \infty$, we conclude that $x_{jr}^{\varepsilon_k}(\overline{\mathbf{w}}_r^{\varepsilon_k}, \mathbf{w}_{-r}^{\varepsilon_k}) \to C_j$ as $k \to \infty$.

Define $\hat{w}_{jr}$ and $\hat{x}_{jr}$ as the limit of $\overline{w}_{jr}^{\varepsilon_k}$ and $x_{jr}^{\varepsilon_k}(\overline{\mathbf{w}}_r^{\varepsilon_k}, \mathbf{w}_{-r}^{\varepsilon_k})$, respectively. From the preceding discussion, as $k \to \infty$ we have the following relations:

$$\hat{w}_{jr} = \lim_{k\to\infty} \overline{w}_{jr}^{\varepsilon^k} = \begin{cases} \overline{w}_{jr}, & \text{if } W_{jr} > 0; \\ 0, & \text{if } W_{jr} = 0. \end{cases} \tag{2.54}$$

$$\hat{x}_{jr} = \lim_{k\to\infty} x_{jr}^{\varepsilon_k}(\overline{\mathbf{w}}_r^{\varepsilon_k}, \mathbf{w}_{-r}^{\varepsilon_k}) = \begin{cases} x_{jr}(\overline{\mathbf{w}}_r, \mathbf{w}_{-r}), & \text{if } W_{jr} > 0; \\ C_j, & \text{if } W_{jr} = 0. \end{cases} \tag{2.55}$$

(2.56)

From (2.54) we conclude $\hat{w}_{jr} \le \overline{w}_{jr}$; and from (2.55) we have $\hat{x}_{jr} \ge x_{jr}(\overline{\boldsymbol{\sigma}}_r, \boldsymbol{\sigma}_{-r})$. But then since the functions d_r and U_r are nondecreasing, we conclude that:

$$\begin{aligned} U_r\big(d_r(\hat{\mathbf{x}}_r)\big) - \sum_j \hat{w}_{jr} &\ge U_r\big(d_r(\mathbf{x}_r(\overline{\boldsymbol{\sigma}}_r, \boldsymbol{\sigma}_{-r}))\big) - \sum_j \overline{w}_{jr} \\ &> U_r\big(d_r(\mathbf{x}_r(\boldsymbol{\sigma}))\big) - \sum_j w_{jr}. \end{aligned}$$

The last inequality follows since $\overline{\mathbf{w}}_r$ is a profitable deviation for user r.

But now recall that the composite function $U_r(d_r(\cdot))$ is continuous in its argument; as a result, from the limits in (2.53), (2.54), and (2.55), we conclude that for sufficiently large k we will have:

$$U_r\big(d_r(\mathbf{x}_r^{\varepsilon_k}(\overline{\mathbf{w}}_r^{\varepsilon_k}, \mathbf{w}_{-r}^{\varepsilon_k}))\big) - \sum_j \overline{w}_{jr}^{\varepsilon_k} > U_r\big(d_r(\mathbf{x}_r^{\varepsilon_k}(\mathbf{w}^{\varepsilon_k}))\big) - \sum_j w_{jr}^{\varepsilon_k}.$$

But this contradicts the fact that $\mathbf{w}^{\varepsilon_k}$ is a Nash equilibrium for the game defined by $(Q_1^{\varepsilon}, \ldots, Q_R^{\varepsilon})$, since we have found a profitable deviation for user r. As a result, no profitable deviation $\overline{\boldsymbol{\sigma}}_r$ can exist for user r in the extended game with respect to the strategy vector $\boldsymbol{\sigma}$; thus we conclude that $\boldsymbol{\sigma}$ is a Nash equilibrium for the extended game, as required. □

The previous theorem demonstrates that the "extended" strategy space eliminates the possibility of the nonexistence of a Nash equilibrium. Indeed, with the extended strategy space, both Examples 2.1 and 2.2 will possess at least one Nash equilibrium.

In Example 2.1, the Nash equilibrium is for the single user to submit a bid of $w = 0$, and to request a rate $\phi = C$. In Example, 2.2, the Nash equilibrium is constructed as follows. First, all users play a single link game for link 1; suppose this results in the Nash equilibrium bid vector $(w_{11}, \ldots, w_{1R})$, with rate allocation to user r given by $x_{1r} = (w_{1r}C_1)/(\sum_s w_{1s})$. We may choose ϕ_{1r} arbitrarily, since it plays no role in the resulting allocation. Suppose each user then submits a bid of $w_{2r} = 0$ to link 2, but requests rate $\phi_{2r} = x_{1r}$ from link 2; since $\sum_r x_{1r} = C_1 < C_2$, these requests will be granted. It is straightforward to check that the strategy vector $(\boldsymbol{\phi}, \mathbf{w})$ is a Nash equilibrium for the extended game. We observe that at this Nash equilibrium, the total payment to link 2 is zero, reflecting the fact that link 2 is not a bottleneck.

We conclude by noting that while Theorem 2.12 establishes existence of a Nash equilibrium in the network case, we have not shown that such a Nash equilibrium is unique. In the special case where $C_j = C$ for all j (all capacities are equal), and each user is identified with exactly one path through the network (fixed routing), it is possible to use an argument analogous to the proof of Theorem 2.2 to show that a Nash equilibrium is unique; in particular, the Nash equilibrium conditions become equivalent to the optimality conditions for a network form of the problem *GAME*. In general, however, such a technique does not apply, and uniqueness of the Nash equilibrium remains an open question.

■ 2.4.2 Efficiency Loss

Let the vector $\boldsymbol{\sigma}$ be a Nash equilibrium of the extended game, i.e., the game defined by $(T_1, \ldots, T_R)$, and let $\mathbf{d}^G = (d_r(\mathbf{x}_r(\boldsymbol{\sigma})), r \in R)$ be the allocation at this Nash equilibrium. Let $\mathbf{d}^S$ denote any optimal solution to the network *SYSTEM* problem. The following theorem demonstrates that the utility lost at any Nash equilibrium is no worse than 25% of the maximum possible aggregate utility, matching the result derived in the single link model. We also note that this result does not require $R > 1$, or U_r to be strictly increasing and continuously differentiable; it is therefore a stronger version of Theorem 2.6 for the single link case.

Theorem 2.13
Suppose that Assumption 2.2 is satisfied. Assume also that $U_r(0) \geq 0$ for all users r. If $\boldsymbol{\sigma}$ is a Nash equilibrium for the extended network game defined by $(T_1, \ldots, T_R)$, and $\mathbf{d}^S$ is any SYSTEM optimal allocation, then:

$$\sum_r U_r\big(d_r(\mathbf{x}_r(\boldsymbol{\sigma}))\big) \geq \frac{3}{4} \sum_r U_r(d_r^S).$$

Proof. For the single user case ($R = 1$), at any Nash equilibrium the single user makes no payment to the link, and is granted any feasible capacity request. Thus any Nash equilibrium allocation yields a rate to user 1 given by $d_1(\mathbf{C})$, where $\mathbf{C}$ is the vector of link capacities. This allocation is an optimal solution to *SYSTEM*, so the theorem is trivially true. We assume without loss of generality, therefore, that $R > 1$ for the remainder of the proof.

The proof consists of three main steps. First, we describe the entire problem in terms of the vector $\mathbf{x}_r(\boldsymbol{\sigma}) = (x_{jr}(\boldsymbol{\sigma}), j \in J)$ of the rate allocations to user r from the network. We show in Lemma 2.14 that Nash equilibria can be characterized in terms of each user r optimally choosing a rate allocation $\overline{\mathbf{x}}_r = (\overline{x}_{jr}, j \in J)$, given the strategies $\boldsymbol{\sigma}_{-r}$ of all other users.

In the second step, we observe that the utility to user r given a vector of rate allocations $\overline{\mathbf{x}}_r$ is exactly $U_r(d_r(\overline{\mathbf{x}}_r))$; we call this a "composite" utility function. In Lemma 2.15, we linearize this composite utility function. Formally, we replace $U_r(d_r(\overline{\mathbf{x}}_r))$ with a linear function $\boldsymbol{\alpha}_r^\top \overline{\mathbf{x}}_r$. The difficulty in this phase of the analysis is that the composite utility function $U_r(d_r(\cdot))$ may not be differentiable, because the max-flow function $d_r(\cdot)$ is not differentiable everywhere; as a result, convex analytic techniques are required.

Finally, we conclude the proof by observing that when the "composite" utility function for user r is linear in the vector of rate allocations $\overline{\mathbf{x}}_r$, the network structure is no longer relevant. In this case the game defined by $(T_1, \dots, T_R)$ decouples into J games, one for each link. We then apply Theorem 2.6 at each link to arrive at the bound in the theorem.

We start by describing the entire problem in terms of the vector $\mathbf{x}_r(\boldsymbol{\sigma}) = (x_{jr}(\boldsymbol{\sigma}), j \in J)$ of the rate allocations to user r from the network. We redefine the problem *SYSTEM* as follows:

$$\text{maximize} \quad \sum_r U_r(d_r(\overline{\mathbf{x}}_r)) \tag{2.57}$$

$$\text{subject to} \quad \sum_r \overline{x}_{jr} \le C_j, \quad j \in J; \tag{2.58}$$

$$\overline{x}_{jr} \ge 0, \quad j \in J,\ r \in R. \tag{2.59}$$

(The notation $\overline{\mathbf{x}}_r$ is used here to distinguish from the function $\mathbf{x}_r(\boldsymbol{\sigma})$.) In this problem, the network only chooses how to allocate rate at each link to the users. The users then solve a max-flow problem to determine the maximum rate at which they can send (this is captured by the function $d_r(\cdot)$). This problem is equivalent to the problem *SYSTEM* as defined in (2.37)-(2.40), because of the definition of $d_r(\cdot)$ in (2.43)-(2.45). We label an optimal solution to this problem by $(\mathbf{x}_r^S, r \in R)$.

Next, we prove a lemma which states that a Nash equilibrium may be characterized in terms of users optimally choosing rate allocations $(\overline{\mathbf{x}}_r, r \in R)$. As before, given a bid vector $\mathbf{w}$, for each link j and each user r we let $W_{jr} = \sum_{s \ne r} w_{js}$. In addition, we define the set $\mathcal{C} \subset \mathbb{R}^J$ by $\mathcal{C} = \{\overline{\mathbf{x}} = (x_j, j \in J) : 0 \le x_j \le C_j\}$. For $\overline{\mathbf{x}}_r \in \mathcal{C}$, we define a function $f_r(\overline{\mathbf{x}}_r; \boldsymbol{\sigma}_{-r})$ as follows:

$$f_r(\overline{\mathbf{x}}_r; \boldsymbol{\sigma}_{-r}) = \begin{cases} -\infty, & \text{if } \overline{x}_{jr} = C_j \text{ for some } j \text{ with } W_{jr} > 0; \\ U_r(d_r(\overline{\mathbf{x}}_r)) - \sum_{j:W_{jr}>0} \frac{W_{jr}\overline{x}_{jr}}{C_j - \overline{x}_{jr}}, & \text{otherwise.} \end{cases} \tag{2.60}$$

Lemma 2.14 *A vector $\boldsymbol{\sigma} = (\boldsymbol{\phi}, \mathbf{w})$ is a Nash equilibrium for the extended game if and only if the following two conditions hold:*

1. *For each link j and each user r, if $W_{jr} = 0$ then $w_{jr} = 0$.*
2. *For each user r:*

$$\mathbf{x}_r(\boldsymbol{\sigma}) \in \arg\max_{\overline{\mathbf{x}}_r \in \mathcal{C}} f_r(\overline{\mathbf{x}}_r; \boldsymbol{\sigma}_{-r}). \tag{2.61}$$

Proof of Lemma. Suppose first that $\boldsymbol{\sigma}$ is a Nash equilibrium. Then consider a link j and user r such that $W_{jr} = 0$. If $w_{jr} > 0$, then user r can achieve exactly the same rate allocation, but lower his total payment, by choosing a bid $\overline{w}_{jr}$ to link j such that $0 < \overline{w}_{jr} < w_{jr}$. This is a profitable deviation, contradicting the assumption that $\boldsymbol{\sigma}$ is a Nash equilibrium. So Condition 1 must hold.

Next, suppose there exists a vector $\overline{\mathbf{x}}_r \in \mathcal{C}$ such that:

$$f_r(\overline{\mathbf{x}}_r; \boldsymbol{\sigma}_{-r}) > f_r(\mathbf{x}_r(\boldsymbol{\sigma}); \boldsymbol{\sigma}_{-r}). \tag{2.62}$$

First, notice that if $W_{jr} > 0$, then the rate allocation rule:

$$x_{jr}(\boldsymbol{\sigma}) = \frac{w_{jr}}{w_{jr} + W_{jr}} C_j$$

implies that:

$$w_{jr} = \frac{W_{jr} x_{jr}(\boldsymbol{\sigma})}{C_j - x_{jr}(\boldsymbol{\sigma})}. \tag{2.63}$$

Since we have already shown $w_{jr} = 0$ if $W_{jr} = 0$, we have:

$$f_r(\mathbf{x}_r(\boldsymbol{\sigma}); \boldsymbol{\sigma}_{-r}) = U_r(d_r(\mathbf{x}_r(\boldsymbol{\sigma}))) - \sum_{j:W_{jr}>0} \frac{W_{jr} x_{jr}(\boldsymbol{\sigma})}{C_j - x_{jr}(\boldsymbol{\sigma})} = T_r(\boldsymbol{\sigma}_r; \boldsymbol{\sigma}_{-r}).$$

On the other hand, consider the following bid vector for user r. If $W_{jr} > 0$, we define:

$$\overline{w}_{jr} = \frac{W_{jr}\overline{x}_{jr}}{C_j - \overline{x}_{jr}}.$$

If $W_{jr} = 0$, then we define $\overline{w}_{jr} = \varepsilon > 0$. We may define $\overline{\phi}_{jr}$ arbitrarily for each link j; it will play no role in user r's payoff.

With the strategy $\overline{\boldsymbol{\sigma}}_r = (\overline{\boldsymbol{\phi}}_r, \overline{\mathbf{w}}_r)$, user r will be allocated a rate $x_{jr}(\overline{\boldsymbol{\sigma}}_r, \boldsymbol{\sigma}_{-r})$ given by:

$$x_{jr}(\overline{\boldsymbol{\sigma}}_r, \boldsymbol{\sigma}_{-r}) = \begin{cases} \overline{x}_{jr}, & \text{if } W_{jr} > 0; \\ C_j, & \text{if } W_{jr} = 0. \end{cases}$$

In particular, we conclude that $x_{jr}(\overline{\boldsymbol{\sigma}}_r, \boldsymbol{\sigma}_{-r}) \geq \overline{x}_{jr}$ for all links j, so that:

$$U_r(d_r(\mathbf{x}_r(\overline{\boldsymbol{\sigma}}_r, \boldsymbol{\sigma}_{-r}))) \geq U_r(d_r(\overline{\mathbf{x}}_r)).$$

The payoff to user r at the strategy vector $\overline{\boldsymbol{\sigma}}_r$ is:

$$\begin{aligned} T_r(\overline{\boldsymbol{\sigma}}_r; \boldsymbol{\sigma}_{-r}) &= U_r(d_r(\mathbf{x}_r(\overline{\boldsymbol{\sigma}}_r, \boldsymbol{\sigma}_{-r}))) - \sum_{j:W_{jr}>0} \overline{w}_{jr} - \sum_{j:W_{jr}=0} \varepsilon \\ &\geq f_r(\overline{\mathbf{x}}_r; \boldsymbol{\sigma}_{-r}) - \sum_{j:W_{jr}=0} \varepsilon. \end{aligned}$$

As a result, for small enough $\varepsilon > 0$ we conclude from (2.62) that $T_r(\overline{\boldsymbol{\sigma}}_r; \boldsymbol{\sigma}_{-r}) > T_r(\boldsymbol{\sigma}_r; \boldsymbol{\sigma}_{-r})$, contradicting the assumption that $\boldsymbol{\sigma}$ was a Nash equilibrium. So Condition 2 must hold as well.

Conversely, suppose that Conditions 1 and 2 of the lemma hold, but that $\boldsymbol{\sigma}$ is not a Nash equilibrium. Fix a user r, and let $\overline{\boldsymbol{\sigma}}_r$ be a profitable deviation for user r. Define $\overline{x}_{jr} = x_{jr}(\overline{\boldsymbol{\sigma}}_r, \boldsymbol{\sigma}_{-r})$ for each link j. Also, observe that if $W_{jr} > 0$, then the relation (2.63) holds, so we have:

$$\begin{aligned} T_r(\overline{\boldsymbol{\sigma}}_r; \boldsymbol{\sigma}_{-r}) &= U_r(d_r(\overline{\mathbf{x}}_r)) - \sum_{j:W_{jr}>0} \frac{W_{jr}\overline{x}_{jr}}{C_j - \overline{x}_{jr}} - \sum_{j:W_{jr}=0} \overline{w}_{jr} \\ &\leq U_r(d_r(\overline{\mathbf{x}}_r)) - \sum_{j:W_{jr}>0} \frac{W_{jr}\overline{x}_{jr}}{C_j - \overline{x}_{jr}} \\ &= f_r(\overline{\mathbf{x}}_r; \boldsymbol{\sigma}_{-r}). \end{aligned}$$

On the other hand, from Condition 1 together with (2.63), we also have:

$$T_r(\boldsymbol{\sigma}_r; \boldsymbol{\sigma}_{-r}) = U_r(d_r(\mathbf{x}_r(\boldsymbol{\sigma}))) - \sum_{j\in r:W_{jr}>0} \frac{W_{jr}x_{jr}(\boldsymbol{\sigma})}{C_j - x_{jr}(\boldsymbol{\sigma})} = f_r(\mathbf{x}_r(\boldsymbol{\sigma}); \boldsymbol{\sigma}_{-r}).$$

Since $\overline{\boldsymbol{\sigma}}_r$ is a profitable deviation for user r, we have $T_r(\overline{\boldsymbol{\sigma}}_r; \boldsymbol{\sigma}_{-r}) > T_r(\boldsymbol{\sigma}_r; \boldsymbol{\sigma}_{-r})$, which implies:

$$f_r(\overline{\mathbf{x}}_r; \boldsymbol{\sigma}_{-r}) > f_r(\mathbf{x}_r(\boldsymbol{\sigma}); \boldsymbol{\sigma}_{-r}).$$

But this violates Condition 2 in the statement of the lemma, a contradiction. So $\boldsymbol{\sigma}$ must have been a Nash equilibrium, as required. □

Now suppose that $\boldsymbol{\sigma}$ is a Nash equilibrium. Our approach is to replace user r by J users (which we call "virtual" users), one at each link j; this process has the effect of *isolating* each of the links, and removes any dependence on network structure. We define the virtual users so that $\boldsymbol{\sigma}$ remains a Nash equilibrium at each single link game. Formally, for each user r, we construct a vector $\boldsymbol{\alpha}_r = (\alpha_{jr}, j \in J)$, and consider a single link game at each link j where user r has linear utility function $U_{jr}(x_{jr}) = \alpha_{jr}x_{jr}$. We choose the vectors $\boldsymbol{\alpha}_r$ so that the Nash equilibrium at each single link game is also given by $\boldsymbol{\sigma}$; we then apply the result of Theorem 2.6 for the single link model to complete the proof of the theorem.

A technical difficulty arises here because the function $U_r(d_r(\cdot))$ may not be differentiable. If the composite function $g_r = U_r(d_r(\cdot))$ were differentiable, then as in the proof of Theorem 2.6, we could find an appropriate vector $\boldsymbol{\alpha}_r$ by choosing $\boldsymbol{\alpha}_r = \nabla g_r(\mathbf{x}_r(\boldsymbol{\sigma}))$. However, in general $U_r(d_r(\cdot))$ is not differentiable; instead, we must choose $\boldsymbol{\alpha}_r$ to be a *supergradient* of $U_r(d_r(\cdot))$, i.e., we require $-\boldsymbol{\alpha}_r$ to be a *subgradient* of $-U_r(d_r(\cdot))$. The reader is referred to the Notation section for reference on these definitions from convex analysis. The key relationship we note is that $\boldsymbol{\gamma}$ is a supergradient of an extended real-valued function $g : \mathbb{R}^J \to \mathbb{R}$ at $\mathbf{x}$ if and only if for all $\overline{\mathbf{x}} \in \mathbb{R}^J$:

$$g(\overline{\mathbf{x}}) \leq g(\mathbf{x}) + \boldsymbol{\gamma}^\top(\overline{\mathbf{x}} - \mathbf{x}).$$

Lemma 2.14 allows us to characterize the Nash equilibrium $\boldsymbol{\sigma}$ as a choice of optimal rate allocation $\overline{\mathbf{x}}_r$ by each user r, given the strategy vector $\boldsymbol{\sigma}_{-r}$ of all other users. We recall the definition of f_r in (2.60); we will now view f_r as an extended real valued function, by defining $f_r(\overline{\mathbf{x}}_r) = -\infty$ for $\overline{\mathbf{x}}_r \notin \mathcal{C}$. We also define extended real-valued functions g_r and h_r on $\mathbb{R}^J$ as follows:

$$g_r(\overline{\mathbf{x}}_r) = \begin{cases} U_r(d_r(\overline{\mathbf{x}}_r)), & \text{if } \overline{\mathbf{x}}_r \in \mathcal{C}; \\ -\infty, & \text{otherwise.} \end{cases}$$

and

$$h_r(\overline{\mathbf{x}}_r; \boldsymbol{\sigma}_{-r}) = \begin{cases} -\infty, & \text{if } \overline{x}_{jr} \geq C_j \text{ for some } j \text{ with } W_{jr} > 0; \\ \\ -\displaystyle\sum_{j: W_{jr} > 0} \frac{W_{jr}\overline{x}_{jr}}{C_j - \overline{x}_{jr}}, & \text{otherwise.} \end{cases}$$

Then we have $f_r = g_r + h_r$ on $\mathbb{R}^J$. We observe that g_r is a concave function of $\overline{\mathbf{x}}_r \in \mathbb{R}^J$. This follows because d_r is a concave function of its argument (as it is the optimal value of the linear program (2.43)-(2.45)), and U_r is nondecreasing and concave. We also note that h_r is a concave function of $\overline{\mathbf{x}}_r \in \mathbb{R}^J$, since $(W_{jr}\overline{x}_{jr})/(C_j - \overline{x}_{jr})$ is a strictly convex function of $\overline{x}_{jr} \in (-\infty, C_j)$ whenever $W_{jr} > 0$. Consequently, f_r is a concave function of $\overline{\mathbf{x}}_r \in \mathbb{R}^J$. Furthermore, the functions f_r, g_r, and h_r are obviously proper—e.g., $g_r(\mathbf{0}) = U_r(0)$, $h_r(\mathbf{0}) = 0$, and $f_r(\mathbf{0}; \boldsymbol{\sigma}_{-r}) = U_r(0)$. We now have the following lemma.

Lemma 2.15 *Let $\boldsymbol{\sigma}$ be a Nash equilibrium. Then for each user r, there exists a vector $\boldsymbol{\alpha}_r = (\alpha_{jr}, j \in J)$ such that:*

1. $\boldsymbol{\alpha}_r \in -\partial[-g_r(\mathbf{x}_r(\boldsymbol{\sigma}))]$.
2. *If $W_{jr} = 0$, then $\alpha_{jr} = 0$.*
3. *If $W_{jr} > 0$, then $\alpha_{jr} > 0$.*
4. *The following relation holds:*

$$\mathbf{x}_r(\boldsymbol{\sigma}) \in \arg\max_{\overline{\mathbf{x}}_r \in \mathcal{C}} \left[\boldsymbol{\alpha}_r^\top \overline{\mathbf{x}}_r - \sum_{j: W_{jr} > 0} \frac{W_{jr}\overline{x}_{jr}}{C_j - \overline{x}_{jr}} \right]. \tag{2.64}$$

Proof of Lemma. Fix a user r. Observe that with the definitions we have made, the domain of g_r is equal to $\mathcal{C}$ (that is, $-\infty < g_r(\overline{\mathbf{x}}_r) < \infty$ for all $\overline{\mathbf{x}}_r \in \mathcal{C}$). Furthermore, for any $\overline{\mathbf{x}}_r$ such that $\overline{x}_{jr} < C_j$ for all j, we have $-\infty < h_r(\overline{\mathbf{x}}_r; \boldsymbol{\sigma}_{-r}) < \infty$. Thus, the relative interior of the domain of $-g_r$ (denoted $\mathrm{ri}(\mathrm{dom}(-g_r))$) has nonempty intersection with the relative interior of the domain of $-h_r$: $\mathrm{ri}(\mathrm{dom}(-g_r)) \cap \mathrm{ri}(\mathrm{dom}(-h_r)) \neq \emptyset$. From Theorem 23.8 in [103], this is is sufficient to ensure that at $\mathbf{x}_r(\boldsymbol{\sigma})$, we have:

$$\partial[-f_r(\mathbf{x}_r(\boldsymbol{\sigma}); \boldsymbol{\sigma}_{-r})] = \partial[-g_r(\mathbf{x}_r(\boldsymbol{\sigma}))] + \partial[-h_r(\mathbf{x}_r(\boldsymbol{\sigma}); \boldsymbol{\sigma}_{-r})]. \tag{2.65}$$

(The summation here of the two subdifferentials on the right hand side is a summation of sets, where $A + B = \{\mathbf{x} + \mathbf{y} : \mathbf{x} \in A, \mathbf{y} \in B\}$; if either A or B is empty, then $A + B$ is empty as well.)

From Condition 2 in Lemma 2.14, we have for all $\overline{\mathbf{x}}_r \in \mathcal{C}$ that:

$$f_r(\mathbf{x}_r(\boldsymbol{\sigma}); \boldsymbol{\sigma}_{-r}) \geq f_r(\overline{\mathbf{x}}_r; \boldsymbol{\sigma}_{-r}).$$

Since $f_r(\overline{\mathbf{x}}_r; \boldsymbol{\sigma}_{-r}) = -\infty$ for $\overline{\mathbf{x}}_r \notin \mathcal{C}$, we conclude $\mathbf{0}$ is a supergradient of f_r at $\mathbf{x}_r(\boldsymbol{\sigma})$, i.e., $\mathbf{0} \in -\partial[-f_r(\mathbf{x}_r(\boldsymbol{\sigma}); \boldsymbol{\sigma}_{-r})]$. As a result, we know from (2.65) that there exists $\boldsymbol{\alpha}_r \in -\partial[-g_r(\mathbf{x}_r(\boldsymbol{\sigma}))]$ and $\boldsymbol{\beta}_r \in -\partial[-h_r(\mathbf{x}_r(\boldsymbol{\sigma}); \boldsymbol{\sigma}_{-r})]$ such that $\boldsymbol{\alpha}_r = -\boldsymbol{\beta}_r$.

We will explicitly compute $\boldsymbol{\beta}_r$. We first note that from Condition 2 of Lemma 2.14, we must have $0 \leq x_{jr}(\boldsymbol{\sigma}) < C_j$ if $W_{jr} > 0$; otherwise the objective function in (2.61) is equal to $-\infty$, which cannot be optimal for user r (e.g., choosing $\overline{\mathbf{x}}_r = \mathbf{0}$ yields an objective function value of $U_r(0) > -\infty$). Now at any point $\overline{\mathbf{x}}_r \in \mathcal{C}$ such that $\overline{x}_{jr} < C_j$ if $W_{jr} > 0$, we note that h_r is in fact differentiable, with:

$$\frac{\partial h_r}{\partial \overline{x}_{jr}}(\overline{\mathbf{x}}_r; \boldsymbol{\sigma}_{-r}) = \begin{cases} -\dfrac{W_{jr}C_j}{(C_j - \overline{x}_{jr})^2}, & \text{if } W_{jr} > 0; \\ 0, & \text{otherwise.} \end{cases}$$

Since h_r is differentiable at $\mathbf{x}_r(\boldsymbol{\sigma})$, we conclude that in fact $-\partial[-h_r(\mathbf{x}_r(\boldsymbol{\sigma}); \boldsymbol{\sigma}_{-r})]$ is a singleton, containing only $\nabla h_r(\mathbf{x}_r(\boldsymbol{\sigma}); \boldsymbol{\sigma}_{-r})$, which is defined by the previous equation. So we must have $\boldsymbol{\beta}_r = \nabla h_r(\mathbf{x}_r(\boldsymbol{\sigma}); \boldsymbol{\sigma}_{-r})$, and thus:

$$\alpha_{jr} = -\beta_{jr} = \begin{cases} \dfrac{W_{jr}C_j}{(C_j - x_{jr}(\boldsymbol{\sigma}))^2}, & \text{if } W_{jr} > 0; \\ 0, & \text{otherwise.} \end{cases}$$

We have established conclusions 1, 2, and 3 of the lemma. To establish conclusion 4, we observe that $\mathbf{0}$ is a supergradient of the following function at $\mathbf{x}_r(\boldsymbol{\sigma})$:

$$\hat{f}_r(\overline{\mathbf{x}}_r; \boldsymbol{\sigma}_{-r}) = \begin{cases} -\infty, & \text{if } \overline{\mathbf{x}}_r \notin \mathcal{C} \\ & \text{or if } \overline{x}_{jr} = C_j \text{ for some } j \text{ with } W_{jr} > 0; \\ \boldsymbol{\alpha}_r^\top \overline{\mathbf{x}}_r - \displaystyle\sum_{j: W_{jr} > 0} \frac{W_{jr}\overline{x}_{jr}}{C_j - \overline{x}_{jr}}, & \text{otherwise.} \end{cases}$$

This observation follows by replacing $g_r(\overline{\mathbf{x}}_r)$ with the following function $\hat{g}_r$ on $\mathbb{R}^J$:

$$\hat{g}_r(\overline{\mathbf{x}}_r) = \begin{cases} \boldsymbol{\alpha}_r^\top \overline{\mathbf{x}}_r, & \text{if } \overline{\mathbf{x}}_r \in \mathcal{C}; \\ -\infty, & \text{otherwise.} \end{cases}$$

Then we have $\hat{f}_r = \hat{g}_r + h_r$; and as before, $\mathrm{ri}(\mathrm{dom}(-\hat{g}_r)) \cap \mathrm{ri}(\mathrm{dom}(-h_r)) \neq \emptyset$, so we have:

$$\partial[-\hat{f}_r(\mathbf{x}_r(\boldsymbol{\sigma}; \boldsymbol{\sigma}_{-r}))] = \partial[-\hat{g}_r(\mathbf{x}_r(\boldsymbol{\sigma}))] + \partial[-h_r(\mathbf{x}_r(\boldsymbol{\sigma}); \boldsymbol{\sigma}_{-r})].$$

The vector $\boldsymbol{\alpha}_r$ is a supergradient of $\hat{g}_r$ for all $\overline{\mathbf{x}}_r \in \mathcal{C}$, i.e., $\boldsymbol{\alpha}_r \in -\partial[-\hat{g}_r(\overline{\mathbf{x}}_r)]$ for all $\overline{\mathbf{x}}_r \in \mathcal{C}$; in particular, $\boldsymbol{\alpha}_r \in -\partial[-\hat{g}(\mathbf{x}_r(\boldsymbol{\sigma}))]$. We have already shown $\{-\boldsymbol{\alpha}_r\} = -\partial[-h(\mathbf{x}_r(\boldsymbol{\sigma}); \boldsymbol{\sigma}_{-r})]$. Thus $\mathbf{0} \in -\partial[-\hat{f}_r(\mathbf{x}_r(\boldsymbol{\sigma}); \boldsymbol{\sigma}_{-r})]$. This implies conclusion 4 of the lemma, as required. □

For each user r, fix the supergradient $\boldsymbol{\alpha}_r$ given by the preceding lemma. We start by observing that for each user r, since $\boldsymbol{\alpha}_r$ is a supergradient of $g_r(\mathbf{x}_r(\boldsymbol{\sigma}))$, we have:

$$U_r(d_r(\mathbf{x}_r^S)) \leq U_r(d_r(\mathbf{x}_r(\boldsymbol{\sigma}))) + \boldsymbol{\alpha}_r^\top(\mathbf{x}_r^S - \mathbf{x}_r(\boldsymbol{\sigma})). \tag{2.66}$$

Now note that if $\boldsymbol{\alpha}_r = 0$ for all r, then we have the following trivial inequality:

$$\sum_r U_r(d_r(\mathbf{x}_r(\boldsymbol{\sigma}))) \geq \sum_r U_r(d_r(\mathbf{x}_r^S)) \geq \frac{3}{4} \sum_r U_r(d_r(\mathbf{x}_r^S)).$$

Thus the theorem holds in this case; so we may assume without loss of generality that $\boldsymbol{\alpha}_r \neq 0$ for at least one user r. This implies that $\alpha_{jr} > 0$ for at least one link j and user r; by the preceding lemma, we must have $W_{jr} > 0$. In particular, we conclude that at least two users are competing for resources at link j.

Since $\alpha_{jr} = 0$ if $W_{jr} = 0$, we have the following simplification of (2.64):

$$\begin{aligned} \mathbf{x}_r(\boldsymbol{\sigma}) &\in \arg \max_{\overline{\mathbf{x}}_r \in \mathcal{C}} \left[\boldsymbol{\alpha}_r^\top \overline{\mathbf{x}}_r - \sum_{j: W_{jr} > 0} \frac{W_{jr} \overline{x}_{jr}}{C_j - \overline{x}_{jr}} \right] \\ &= \arg \max_{\overline{\mathbf{x}}_r \in \mathcal{C}} \left[\sum_{j: W_{jr} > 0} \left(\alpha_{jr} \overline{x}_{jr} - \frac{W_{jr} \overline{x}_{jr}}{C_j - \overline{x}_{jr}} \right) \right]. \end{aligned}$$

The maximum on the right hand side of the preceding expression decomposes into separate maximizations for each link j with $W_{jr} > 0$. We conclude that for each link j with $W_{jr} > 0$, we in fact have:

$$x_{jr}(\boldsymbol{\sigma}) \in \arg \max_{0 \leq \overline{x}_{jr} \leq C_j} \left[\alpha_{jr} \overline{x}_{jr} - \frac{W_{jr} \overline{x}_{jr}}{C_j - \overline{x}_{jr}} \right].$$

Fix now a link j with $\sum_r w_{jr} > 0$. We view the users as playing a single link game at link j, with utility function for user r given by $U_{jr}(x_{jr}) = \alpha_{jr} x_{jr}$. The preceding expression states that Condition 2 of Lemma 2.14 is satisfied. Furthermore, since $\sum_r w_{jr} > 0$ and $\boldsymbol{\sigma}$ is a Nash equilibrium for the network game, from Condition 1 in Lemma 2.14 there must exist at least two users r_1, r_2 such that $W_{jr_1}, W_{jr_2} > 0$, so in particular, $W_{jr} > 0$ for all users r. Thus Condition 1 of Lemma 2.14 is vacuously satisfied for the single link game; and we conclude that $\boldsymbol{\sigma}$ is a Nash equilibrium for this single link game at link j. More precisely, we have that $(w_{j1}, \ldots, w_{jR})$ is a Nash equilibrium for the single link game at link j, when R users with utility functions $(U_{j1}, \ldots, U_{jR})$ compete for link j. Since $W_{jr} > 0$ for all r, we know $\alpha_{jr} > 0$ for all users r from the preceding lemma, so U_{jr} is strictly increasing for each r; and since $R > 1$, we apply Theorem 2.6 to conclude that:

$$\sum_r \alpha_{jr} x_{jr}(\boldsymbol{\sigma}) \geq \frac{3}{4} \left(\max_r \alpha_{jr} \right) C_j. \tag{2.67}$$

(The right hand side is $3/4$ of the optimal value of *SYSTEM* for a single link of capacity C_j, when each user r has linear utility $U_r(x_{jr}) = \alpha_{jr} x_{jr}$.)

We now complete the proof of the theorem, by following the proof of Lemma 2.7. Note that since $W_{jr} = 0$ implies $\alpha_{jr} = 0$ from Lemma 2.15, the following relation holds:

$$\sum_r \sum_{j:W_{jr}>0} \alpha_{jr} x_{jr}^S = \sum_j \sum_r \alpha_{jr} x_{jr}^S.$$

Thus we have:

$$\sum_r \boldsymbol{\alpha}_r^\top \mathbf{x}_r^S = \sum_r \sum_{j:W_{jr}>0} \alpha_{jr} x_{jr}^S = \sum_j \sum_r \alpha_{jr} x_{jr}^S \le \sum_j \left(\max_r \alpha_{jr}\right) C_j. \tag{2.68}$$

Now note that we can assume that $\sum_r U_r(d_r(\mathbf{x}_r^S)) > 0$. If instead $\sum_r U_r(d_r(\mathbf{x}_r^S)) = 0$, then it must be the case that $\sum_r U_r(d_r(\mathbf{x}_r(\boldsymbol{\sigma}))) = 0$, since $\mathbf{x}^S$ is an optimal solution to *SYSTEM* and all utility functions are nonnegative. Thus the result of the theorem trivially holds in this case; so we assume without loss of generality that $\sum_r U_r(d_r(\mathbf{x}_r^S)) > 0$. Given this fact, we reason as follows, using (2.66) for the first inequality, and (2.68) for the second:

$$\begin{aligned}
\frac{\sum_r U_r(d_r(\mathbf{x}_r(\boldsymbol{\sigma})))}{\sum_r U_r(d_r(\mathbf{x}_r^S))} &\ge \frac{\sum_r \left(U_r(d_r(\mathbf{x}_r(\boldsymbol{\sigma}))) - \boldsymbol{\alpha}_r^\top \mathbf{x}_r(\boldsymbol{\sigma})\right) + \sum_r \boldsymbol{\alpha}_r^\top \mathbf{x}_r(\boldsymbol{\sigma})}{\sum_r \left(U_r(d_r(\mathbf{x}_r(\boldsymbol{\sigma}))) + \boldsymbol{\alpha}_r^\top (\mathbf{x}_r^S - \mathbf{x}_r(\boldsymbol{\sigma}))\right)} \\
&= \frac{\sum_r \left(U_r(d_r(\mathbf{x}_r(\boldsymbol{\sigma}))) - \boldsymbol{\alpha}_r^\top \mathbf{x}_r(\boldsymbol{\sigma})\right) + \sum_r \boldsymbol{\alpha}_r^\top \mathbf{x}_r(\boldsymbol{\sigma})}{\sum_r \left(U_r(d_r(\mathbf{x}_r(\boldsymbol{\sigma}))) - \boldsymbol{\alpha}_r^\top \mathbf{x}_r(\boldsymbol{\sigma})\right) + \sum_r \boldsymbol{\alpha}_r^\top \mathbf{x}_r^S} \\
&\ge \frac{\sum_r \left(U_r(d_r(\mathbf{x}_r(\boldsymbol{\sigma}))) - \boldsymbol{\alpha}_r^\top \mathbf{x}_r(\boldsymbol{\sigma})\right) + \sum_j \sum_r \alpha_{jr} x_{jr}(\boldsymbol{\sigma})}{\sum_r \left(U_r(d_r(\mathbf{x}_r(\boldsymbol{\sigma}))) - \boldsymbol{\alpha}_r^\top \mathbf{x}_r(\boldsymbol{\sigma})\right) + \sum_j \left(\max_r \alpha_{jr}\right) C_j}.
\end{aligned} \tag{2.69}$$

Since $U_r(d_r(\mathbf{0})) = U_r(0) = 0$, applying the fact that $\boldsymbol{\alpha}_r$ is a supergradient we have:

$$U_r(d_r(\mathbf{x}_r(\boldsymbol{\sigma}))) - \boldsymbol{\alpha}_r^\top \mathbf{x}_r(\boldsymbol{\sigma}) \ge 0.$$

We also have:

$$0 \le \sum_j \sum_r \alpha_{jr} x_{jr}(\boldsymbol{\sigma}) \le \sum_j \left(\max_r \alpha_{jr}\right) C_j.$$

So we conclude from relations (2.67) and (2.69) that:

$$\frac{\sum_r U_r(d_r(\mathbf{x}_r(\boldsymbol{\sigma})))}{\sum_r U_r(d_r(\mathbf{x}_r^S))} \ge \frac{\sum_j \sum_r \alpha_{jr} x_{jr}(\boldsymbol{\sigma})}{\sum_j \left(\max_r \alpha_{jr}\right) C_j} \ge \frac{3}{4}.$$

Observe that all denominators in this chain of inequalities are nonzero, since $\boldsymbol{\alpha}_r \neq 0$ for at least one user r implies that:

$$\sum_j (\max_r \alpha_{jr}) C_j > 0.$$

Since $\boldsymbol{\sigma}$ was assumed to be a Nash equilibrium, this completes the proof of the theorem. □

The preceding theorem uses the single link result to determine the worst case efficiency loss for general networks. Note that since we knew from Theorem 2.6 that the bound of $3/4$ was essentially tight for single link games, and a single link is a special case of a general network, the $3/4$ bound is also tight in this setting. In particular, note that a single link yields the worst efficiency loss. This is similar to a result observed by Roughgarden for traffic routing games [106], where the worst efficiency loss occurs in very simple networks.

■ 2.4.3 A Comparison to Proportionally Fair Pricing

Our motivation for considering the network game described in this section comes from the network model described by Kelly in [62]. Rather than a model where users submit individual bids to each link in the network where they desire service, as we have considered, Kelly described a market mechanism known as *proportionally fair pricing* where each user r only submits a *total* payment w_r, and receives in return a rate allocation $d_r^P(\mathbf{w})$ (where $\mathbf{w}$ is the composite vector of total payments from all users). The network chooses the splitting of w_r into individual payments w_{jr} to each link j; and in turn, the network determines the aggregation of individual allocations $(x_{jr}, j \in J)$ into a rate allocation d_r^P to user r.

The simplest development of this network allocation model is axiomatic. We define the *proportionally fair pricing axioms* as follows.

Definition 2.1
Given $\mathbf{w} \geq 0$*, a vector* $\mathbf{d} \geq 0$ *satisfies the* proportionally fair pricing axioms *if there exist vectors* $\boldsymbol{\mu} \geq 0$*,* $\boldsymbol{\lambda} \geq 0$*, and* $\mathbf{y} \geq 0$ *such that:*

1. *For all* r*,* $d_r = \sum_{p \in r} y_p$*, and* $\lambda_r = \min_{p \in r} \sum_{j \in p} \mu_j$.
2. *If* $w_r = 0$*, then* $d_r = 0$.
3. *If* $w_r > 0$*, then* $w_r = \lambda_r d_r$.
4. *If* $y_p > 0$*, then* $\lambda_r = \sum_{j \in p} \mu_j$.
5. *If* $\sum_{p: j \in p} y_p < C_j$*, then* $\mu_j = 0$.

We interpret μ_j as the price set at link j; λ_r as the price per unit rate experienced by user r; y_p as the rate allocated to path p; and d_r as the rate allocation to user r. We interpret these axioms as follows. The first axiom defines d_r as the total rate allocation to user r, summed over all paths $p \in r$; and λ_r as the price seen by user r, determined as the aggregate price of the cheapest path(s) available to user r. The second axiom requires that if a user bids zero, he is allocated zero rate in return. The third axiom ensures that if a user r submits a positive bid, his allocation is equal to the bid divided by his price λ_r; thus w_r is the total payment made by user r. The fourth axiom states that the only paths on which user r sends positive rate are those with the lowest aggregate price among available paths. Finally, the last axiom is a complementary slackness condition, which ensures a positive link price is set only if the link is saturated.

The primary restrictions on this pricing mechanism are the use of w_r to represent the total payment made by user r to the network, and the fact that the network sets a single price of μ_j at each link j. Kelly shows in [62] that this uniquely defines the allocation to user r.

Proposition 2.16 (Kelly [62])
Given a vector $\mathbf{w} \geq 0$, there exists a unique allocation $\mathbf{d}^P(\mathbf{w})$ satisfying the proportionally fair pricing axioms. This allocation is the unique optimal solution to the following optimization problem:

$$\begin{aligned} \textit{maximize} \quad & \sum_{r: w_r > 0} w_r \log d_r \\ \textit{subject to} \quad & \mathbf{A}\mathbf{y} \leq \mathbf{C}; \\ & \mathbf{H}\mathbf{y} = \mathbf{d}; \\ & d_r = 0, \ \ \textit{if } w_r = 0; \\ & y_p \geq 0, \ \ p \in P. \end{aligned}$$

Note that formally, for fixed $\mathbf{w}$ the allocation $\mathbf{d}^P(\mathbf{w})$ is an optimal solution to an analogue of *SYSTEM* with the utility of user r equal to $U_r(d_r) = w_r \log d_r$. Several points are straightforward to check. First, in a single link setting, this mechanism is exactly the mechanism studied in Section 2.1. Second, if users are price taking, Kelly shows in [62] that there exists a competitive equilibrium, and the resulting allocation is an optimal solution to *SYSTEM*.

We expect that proportionally fair pricing by the network will ensure the total bid w_r of user r is split among the links to maximize the rate delivered to user r, given the splitting of the remaining users' bids among the links. This follows from the form of the objective function $\sum_{r: w_r > 0} w_r \log d_r$: if $w_r > 0$, then the term $w_r \log d_r$ is maximized by "purchasing" the maximum rate d_r possible for user r, given the splitting across links of the payments w_s by users $s \neq r$. This observation suggests that at a Nash equilibrium $\boldsymbol{\sigma}$ of the network game presented in Section 2.4.1, the resulting allocation will be the same as $d_r^P(w_1, \ldots, w_R)$, where w_r is the total payment of user r. This is formalized in the following proposition.

Proposition 2.17
Suppose that Assumption 2.2 is satisfied. Let $\boldsymbol{\sigma}$ be a Nash equilibrium of the extended game defined by $(T_1, \ldots, T_R)$, and let $d_r(\mathbf{x}_r(\boldsymbol{\sigma}))$ be the resulting allocation to user r. Let $w_r = \sum_j w_{jr}$. Then:

1. *If $w_r > 0$, then $d_r(\mathbf{x}_r(\boldsymbol{\sigma})) = d_r^P(w_1, \ldots, w_R)$.*
2. *If $w_r = 0$, then $d_r(\mathbf{x}_r(\boldsymbol{\sigma})) \geq d_r^P(w_1, \ldots, w_R) = 0$.*

Proof. Let $d_r = d_r(\mathbf{x}_r(\boldsymbol{\sigma}))$ if $w_r > 0$, and $d_r = 0$ if $w_r = 0$. We demonstrate the existence of $\boldsymbol{\mu}$, $\boldsymbol{\lambda}$, and $\mathbf{y}$ such that the proportionally fair pricing axioms are satisfied for $\mathbf{d}$. For each j, let $\mu_j = (\sum_r w_{jr})/C_j$. For each r, let $\lambda_r = \min_{p \in r} \sum_{j \in p} \mu_j$. If $w_r > 0$, then let $\mathbf{y}_r = (y_p, p \in r)$ be an optimal solution to (2.43)-(2.45) when $\overline{\mathbf{x}}_r = \mathbf{x}_r(\boldsymbol{\sigma})$; otherwise, if $w_r = 0$, let $\mathbf{y}_r = \mathbf{0}$. With these definitions, we have $d_r = \sum_{p \in r} y_p$, so the first and second axioms are trivially satisfied.

Now suppose that for some j, we have $\sum_{p:j \in p} y_p < C_j$ while $\mu_j > 0$. For any user r with $w_{jr} > 0$, we must have $\sum_{p \in r: j \in p} y_p = x_{jr}(\boldsymbol{\sigma})$; otherwise user r can profitably deviate by reducing w_{jr}. For any user r with $w_{jr} = 0$, we have $x_{jr}(\boldsymbol{\sigma}) = 0 = \sum_{p \in r: j \in p} y_p$. Thus we have $\sum_{p:j \in p} y_p = \sum_r x_{jr}(\boldsymbol{\sigma}) < C_j$, while $\mu_j > 0$—a contradiction to the definition of the pricing mechanism in Section 2.4.1. Thus the fifth axiom is satisfied.

Now suppose that $y_p > 0$, but that $\sum_{j \in p} \mu_j > \lambda_r$; it then follows that $\sum_{j \in p} w_{jr} > 0$. Choose $p' \in r$ such that $\sum_{j \in p'} \mu_j = \lambda_r$. Then it is straightforward to check that a profitable deviation for user r is to infinitesimally reduce $\sum_{j \in p} w_{jr}$ and y_p, and infinitesimally increase $\sum_{j \in p'} w_{jr}$ and $y_{p'}$, such that the aggregate rate allocated to user r remains unchanged, but the total payment to the network made by user r decreases. Thus $\boldsymbol{\sigma}$ could not have been a Nash equilibrium; we conclude that the fourth axiom must hold.

Finally, to verify the third axiom, note that if $w_{jr} > 0$, then the constraint (2.44) must be binding for link j at the solution $\mathbf{y}_r$; otherwise, user r could profitably deviate by lowering w_{jr}. Furthermore, if $w_{jr} > 0$, then we have $x_{jr}(\boldsymbol{\sigma}) = w_{jr}/\mu_j$. Using these facts and the fourth axiom, we argue as follows:

$$\sum_j w_{jr} = \sum_{p \in r} y_p \sum_{j \in p} \mu_j = \sum_{p \in r} y_p \lambda_r = \lambda_r d_r(\mathbf{x}_r(\boldsymbol{\sigma})).$$

This establishes the third axiom, completing the proof. □

Thus, at a Nash equilibrium of the game studied here, the allocation is nearly the same as that made in the network model studied by Kelly in [62]; the only difference is that the extended game of Section 2.4.1 allows a user to receive a positive rate even if he does not submit a positive bid, whereas the proportionally fair pricing axioms require that such a user receives no rate allocation (cf. Axiom 2 of Definition 2.1). We may interpret this result as an insight into the information structure of our game. By definition, at a Nash equilibrium of the game defined by $(T_1, \dots, T_R)$ (see (2.51)), the users have chosen their individual payments to each link optimally, while keeping the payments by other users to each link fixed. However, given the vector $\boldsymbol{\sigma}_{-r}$ of other strategies, it suffices for user r to choose only his total payment w_r to the network; given this total payment and the strategy vector $\boldsymbol{\sigma}_{-r}$, the splitting of the payment w_r across the links should be done so as to maximize the rate allocated to user r. From this argument, we see that at a Nash equilibrium of the game defined by $(T_1, \dots, T_R)$, it is as if the users have chosen their *total payment* to the network optimally, while keeping the payments by other users *to each link* fixed; and, from Proposition 2.17, the allocation to any user r with $w_r > 0$ is then made according to the network model described in [62]. Thus we may view such an equilibrium as a *limited response equilibrium* of the game where users submit total payments to the network: users anticipate the effect of a change in their total payment on the prices of the links, but not on the reallocation of other users' total payments among the links. An alternative model, requiring greater deductive capability on the part of the users, would require that each user r completely anticipate the effect of changing his total payment w_r on the reallocation of the other users' total payments $(w_1, \dots, w_{r-1}, w_{r+1}, \dots, w_R)$ among the links of the network. Hajek and Yang have analyzed this game, and show that in general Nash equilibria may not exist; and further, even when Nash equilibria do exist, the efficiency loss may be arbitrarily high, in contrast to the result of Theorem 2.13 [53].

2.5 Extensions

The next two sections extend the model of this chapter to cover the possibility of stochastic capacity (Section 2.5.1), and more general resource allocation environments (Section 2.5.2).

2.5.1 Stochastic Capacity

In this section we consider a model where supply is stochastic, rather than predetermined. We begin by considering a single link model. Let the capacity of the single link be $C > 0$, with distribution $\mathbb{P}$. We assume that $0 < \mathbb{E}[C] = \int_0^\infty C d\mathbb{P}(C) < \infty$. We also assume, as in Section 2.1, that R users share the link. User r has a utility function U_r, and the utilities U_r are assumed to satisfy Assumption 2.1.

In this setting, we will define an allocation in terms of the *fractions* allocated to each user, rather than the absolute amount of resource allocated. Formally, we define the problem *SYSTEM* as:

SYSTEM:

$$\text{maximize} \quad \sum_r \mathbb{E}[U_r(\pi_r C)] \tag{2.70}$$

$$\text{subject to} \quad \sum_r \pi_r = 1; \tag{2.71}$$

$$\pi_r \geq 0, \quad r \in R. \tag{2.72}$$

Notice that this problem chooses the fractions π_r allocated to each resource optimally *ex ante*; that is, before the true supply has been realized.

Our key insight in analyzing this model is that stochastic capacity is equivalent to a model with deterministic capacity $C = 1$, for an appropriate choice of utility functions. Formally, for each user r, define $\overline{U}_r$ as follows:

$$\overline{U}_r(\pi_r) = \mathbb{E}[U_r(\pi_r C)]. \tag{2.73}$$

We have the following proposition.

Proposition 2.18
Suppose that Assumption 2.1 is satisfied by the utility functions $U_1, \ldots, U_R$. Then Assumption 2.1 is also satisfied by the utility functions $\overline{U}_1, \ldots, \overline{U}_R$.

Proof. We first show $\overline{U}_r$ is continuous and continuously differentiable. Let $\pi_r^n \to \pi_r$ as $n \to \infty$, where $\pi_r^n \geq 0$. Then $U_r(\pi_r^n C) \to U_r(\pi_r C)$ for all $C > 0$. Because U_r is strictly increasing and concave, we have:

$$U_r(0) \leq U_r(\pi_r^n C) \leq U_r(0) + U_r'(0)\pi_r^n C \leq U_r(0) + U_r'(0)C.$$

Since C is integrable, we may apply the dominated convergence theorem to conclude that $\mathbb{E}[U_r(\pi_r^n C)] \to \mathbb{E}[U_r(\pi_r C)]$ as $n \to \infty$. Thus $\overline{U}_r$ is continuous. Since U_r is concave and satisfies Assumption 2.1, it has bounded derivative, and:

$$0 \leq \frac{U_r(\pi_r^n C) - U_r(\pi_r C)}{\pi_r^n - \pi_r} \leq U_r'(0)C.$$

Thus again applying the dominated convergence theorem we conclude that $\overline{U}_r$ is differentiable, with derivative $\overline{U}_r'(\pi_r) = \mathbb{E}[U_r'(\pi_r C)C]$. Furthermore, the fact that U_r' is bounded allows us to use the dominated convergence theorem once more to show that $\overline{U}_r'$ is continuous as well.

It remains to be shown that $\overline{U}_r$ is concave and strictly increasing. Concavity follows immediately from concavity of U_r, because expectation is linear; i.e., if $0 < \delta < 1$ and

$\pi_r^1, \pi_r^2 > 0$, then for all $C > 0$:

$$U_r(\delta\pi_r^1 C + (1-\delta)\pi_r^2 C) \geq \delta U_r(\pi_r^1 C) + (1-\delta)U_r(\pi_r^2 C),$$

and taking expectations yields concavity of $\overline{U}_r$. Finally, since U_r is strictly increasing, if $\pi_r^1 > \pi_r^2$ then we have $U_r(\pi_r^1 C) > U_r(\pi_r^2 C)$ for for all $C > 0$. Taking expectations shows that $\overline{U}_r$ is strictly increasing (since $\mathbb{E}[C] > 0$). □

The preceding lemma allows us to extend the main results of Sections 2.1 and 2.2 to the setting of stochastic capacity. We start with the following proposition.

Proposition 2.19
Suppose that Assumption 2.1 is satisfied. Then there exists a vector $\boldsymbol{\pi}^S$ *solving* (2.70)-(2.72). *Furthermore,* $\boldsymbol{\pi}^S$ *is an optimal solution to* (2.70)-(2.72) *if and only if* $\boldsymbol{\pi}^S$ *is an optimal solution to* (2.37)-(2.39) *with utility functions* $\overline{U}_1, \ldots, \overline{U}_R$ *and capacity* $C = 1$.

Proof. From Proposition 2.18, the objective function (2.70) is continuous and the feasible region (2.71)-(2.72) is compact. Furthermore, under the identification (2.73), the problem (2.70)-(2.72) becomes equivalent to (2.37)-(2.39). □

We continue to use the same pricing mechanism even in the presence of stochastic capacity. Each user submits a bid w_r, and receives a fraction $w_r/(\sum_s w_s)$ of the realized capacity C. Formally, the payoff to user r is given by:

$$\overline{Q}_r(w_r; \mathbf{w}_{-r}) = \begin{cases} \mathbb{E}\left[U_r\left(\dfrac{w_r}{\sum_s w_s} C\right)\right] - w_r, & \text{if } w_r > 0; \\ \\ U_r(0), & \text{if } w_r = 0. \end{cases} \tag{2.74}$$

Now observe that $\overline{Q}_r(w_r; \mathbf{w}_{-r})$ is identical to the payoff $Q_r(w_r; \mathbf{w}_{-r})$ if we substitute the utility function $\overline{U}_r$ and capacity $C = 1$ in the definition (2.13). This observation leads to the following proposition.

Proposition 2.20
Assume that $R > 1$, *and suppose that Assumption 2.1 is satisfied. Then there exists a Nash equilibrium* $\mathbf{w} \geq 0$ *for the game defined by* $(\overline{Q}_1, \ldots, \overline{Q}_R)$. *Furthermore,* $\mathbf{w} \geq 0$ *is a Nash equilibrium for the game defined by* $(\overline{Q}_1, \ldots, \overline{Q}_R)$ *if and only if* $\mathbf{w}$ *is a Nash equilibrium for the game defined by* $(Q_1, \ldots, Q_R)$ *when the utility function of each user* r *is* $\overline{U}_r$ *and the capacity is* $C = 1$.

The next proposition shows that the aggregate utility at a Nash equilibrium is no worse than 75% of the aggregate utility at an optimal solution to *SYSTEM*. The intuition is clear: we simply apply Theorem 2.6 to a game where the utility function of each user r is $\overline{U}_r$, and the capacity is $C = 1$.

Theorem 2.21
Suppose that Assumption 2.1 is satisfied. Suppose also that $U_r(0) \geq 0$ for all r. Let $\mathbf{w}$ be a Nash equilibrium of the game defined by $(\overline{Q}_1, \ldots, \overline{Q}_R)$, and define:

$$\pi_r^G = \frac{w_r}{\sum_s w_s}.$$

If $\boldsymbol{\pi}^S$ is any optimal solution to SYSTEM, then:

$$\sum_r \mathbb{E}[U_r(\pi_r^G C)] \geq \frac{3}{4} \sum_r \mathbb{E}[U_r(\pi_r^S C)].$$

Proof. We only need to check that $U_r(0) \geq 0$ implies that $\overline{U}_r(0) \geq 0$; but this is trivial from the definition of $\overline{U}_r$. Thus $\overline{U}_r(0) \geq 0$ and $\overline{U}_1, \ldots, \overline{U}_R$ satisfy Assumption 2.1 (from Proposition 2.18). Thus by applying Propositions 2.19 and 2.20 together with Theorem 2.6, the result follows. □

It is straightforward to extend this analysis to a network context, using exactly the same methods as we applied in Section 2.4. We include this extension in our general resource allocation model in the following section.

■ 2.5.2 A General Resource Allocation Game

In this section we consider an extension to more general resource allocation games. Suppose that there are J infinitely divisible scarce resources, and R users require these resources. As before, let C_j be the total available amount of resource j; we assume that C_j is stochastic, and the joint distribution of $C_1, \ldots, C_J$ is given by $\mathbb{P}$. We also require that $\mathbb{P}(C_j = 0) = 0$ for all j, and $\mathbb{E}[C_j] < \infty$ for all j. We let x_{jr} denote the amount of resource j allocated to user r. The key property which drives the model of this section is the assumption that user r receives a utility $V_r(\mathbf{x}_r)$ from the allocation $\mathbf{x}_r = (x_{jr}, j \in J)$, where V_r satisfies the following assumption.

Assumption 2.3
For each r, the utility function $V_r(\mathbf{x}_r)$ is a concave and continuous function of the vector $\mathbf{x}_r \geq 0$. In addition, V_r is nondecreasing*; that is, if $x_{jr} \geq \overline{x}_{jr}$ for all $j \in J$, then $V_r(\mathbf{x}_r) \geq V_r(\overline{\mathbf{x}}_r)$.*

Of course, one example where these conditions are satisfied is given by the model of this chapter, where the resources represent links in a communication network, and each user requires a subset of these resources. User r receives a nondecreasing, concave, continuous utility $U_r(d_r)$ as a function of the total rate d_r obtained from the network; and the rate $d_r(\mathbf{x}_r)$ is determined by solving the max-flow problem (2.43)-(2.45).

In this case, the composite function $U_r(d_r(\mathbf{x}_r))$ is concave and nondecreasing in the argument $\mathbf{x}_r$.

Another example may be described by interpreting each resource j as a distinct raw material, and $V_r(\mathbf{x}_r)$ as the profits of a firm r which has access to x_{jr} units of raw material j for each $j \in J$. In this case, the assumption that V_r is concave corresponds to decreasing marginal returns; and the assumption that V_r is nondecreasing implies profits should not fall as the raw materials available increase.

We suppose now that the users play a game to acquire resources as described in Section 2.4.1. However, because resource capacities are stochastic, we rewrite the game in terms of the *fractions* of each resource allocated to user r. In particular, each user r chooses a requested fractional resource allocation ϕ_{jr} and makes a bid w_{jr} to each resource $j \in J$. Given the composite strategy vector $\boldsymbol{\sigma} = (\boldsymbol{\phi}, \mathbf{w})$, resource j then allocates a fraction $\pi_{jr}(\boldsymbol{\sigma})$ to user r, where $\pi_{jr}(\cdot)$ is defined as follows:

1. If $\sum_s w_{js} > 0$, then:
$$\pi_{jr}(\boldsymbol{\sigma}) = \frac{w_{jr}}{\sum_s w_{js}}. \tag{2.75}$$

2. If $\sum_s w_{js} = 0$, and $\sum_s \phi_{js} \leq 1$, then:
$$\pi_{jr}(\boldsymbol{\sigma}) = \phi_{jr}. \tag{2.76}$$

3. If $\sum_s w_{js} = 0$, and $\sum_s \phi_{js} > 1$, then:
$$\pi_{jr}(\boldsymbol{\sigma}) = 0. \tag{2.77}$$

Once the capacities C_j are realized, user r receives an allocation $x_{jr}(\boldsymbol{\sigma}) = \pi_{jr}(\boldsymbol{\sigma})C_j$ from resource j. The expected payoff to user r is:

$$Y_r(\boldsymbol{\sigma}_r; \boldsymbol{\sigma}_{-r}) = \mathbb{E}[V_r(\mathbf{x}_r(\boldsymbol{\sigma}))] - \sum_j w_{jr}.$$

Following the proofs of Theorem 2.12 and Proposition 2.20, but replacing $U_r(d_r(\cdot))$ with $\mathbb{E}[V_r(\cdot)]$ for each r, we may prove the following theorem.

Theorem 2.22
Suppose that Assumption 2.3 is satisfied. Then a Nash equilibrium exists for the game defined by $(Y_1, \ldots, Y_R)$.

More importantly, we would like to compare the performance at any Nash equilibrium of this game with an "efficient" allocation. As in the preceding development, we define the problem *SYSTEM* as follows:

SYSTEM:

$$\text{maximize} \quad \sum_r \mathbb{E}[V_r(\overline{\pi}_{1r}C_1, \dots, \overline{\pi}_{Jr}C_J)] \tag{2.78}$$

$$\text{subject to} \quad \sum_r \overline{\pi}_{jr} \leq 1, \quad j \in J; \tag{2.79}$$

$$\overline{\pi}_{jr} \geq 0, \quad j \in J,\ r \in R. \tag{2.80}$$

Since the objective function is continuous and the feasible region is compact, an optimal solution exists for this problem. Again, following the proof of Theorem 2.13 together with Theorem 2.21, we may prove the following result.

Theorem 2.23
Suppose that Assumption 2.3 is satisfied. Assume also that $V_r(\mathbf{0}) \geq 0$ for all users r. Let σ be a Nash equilibrium of the game defined by $(Y_1, \dots, Y_R)$. Let $\boldsymbol{\pi}^S = (\boldsymbol{\pi}_r^S, r \in R)$ be any optimal solution to SYSTEM, and let $x_{jr}^S = \pi_{jr}^S C_j$. Then:

$$\sum_r \mathbb{E}[V_r(\mathbf{x}_r(\boldsymbol{\sigma}))] \geq \frac{3}{4} \sum_r \mathbb{E}[V_r(\mathbf{x}_r^S)].$$

The preceding theorem shows that the essential structure in the network context is the bidding scheme which allows each resource to operate its own "market." Each user then decides how to employ allocated resources, resulting in the expected utility $\mathbb{E}[V_r(\mathbf{x}_r(\boldsymbol{\sigma}))]$. This decoupling between the pricing mechanism employed at each resource and the eventual use of the resources by the end users allows the extension of the result of Theorem 2.6 from a single resource context to a general multiple resource context.

■ 2.6 Chapter Summary

This chapter has considered a model of resource allocation in settings where the supply of the resource available is inelastic—i.e., it does not respond to price. We have shown that when users are price anticipating, the efficiency loss is no more than 25% (Theorem 2.6); and further, this basic result extends to a setting of general networks (Theorem 2.13), as well as to a setting of stochastic demand (Theorem 2.21). The main motivation for such a model comes from usage based pricing in communication networks [62]. We close this chapter with a closer examination of the applicability of this pricing model.

It is unrealistic to expect end users to pay for network rate allocation on a real time basis. Rather, as advocated by Key [66] (and, in a different context, by Gibbens and Kelly [44]), a more feasible scenario is that *brokers* bid for network resources on behalf of users, and in turn offer longer term contracts to users. In this context, it is both more likely to expect such a pricing scheme to be implemented, and also more likely that brokers will have sufficient market power to anticipate the effect of their bids on prices. Such a scheme would admit the benefits of a usage based pricing scheme in terms of network efficiency, while being feasible to implement and maintain.

Once we consider such a domain, however, we must acknowledge the great importance of network provider competition in future models of network pricing. Indeed, although the results of this chapter shed insight into the efficiency properties of simple market-clearing mechanisms, it is less clear whether such market mechanisms will ever be implemented in the decentralized, deregulated world of the current Internet. Thus understanding how market pressures will affect the pricing strategies of competing Internet providers remains a critical research issue for future communication networks. We refer the reader to Odlyzko's insightful critique of the issues involved in this dimension [95].

CHAPTER 3

Multiple Consumers, Elastic Supply

In the previous chapter, we developed market mechanisms for a setting where the available supply of resources is inelastic. We now turn our attention to analyzing market mechanisms for a setting where supply is elastic, i.e., where supply can vary with price. In the models we discuss in this chapter, we will replace the fixed capacity C of the model of Section 2.1 with a *cost function* $C(f)$, which gives the monetary cost incurred if the link manager allocates f units of data rate to the users. With this modification, we wish to design mechanisms which maximize the aggregate utility of the users less the cost incurred at the link, i.e., the aggregate surplus.

We will start by investigating a market mechanism which is the natural extension of the price mechanism considered in Section 2.1 to a setting with elastic supply. The mechanism we describe was first considered by Kelly et al. in [65] (motivated by the proposal made in [62]). In the special case of a single link, the mechanism works as follows. Each user submits a bid, or total *willingness-to-pay*, to the link manager. This represents the total amount the user expects to pay. The link manager then chooses both a total rate and a price such that the product of price and rate is equal to the total revenue, and the price is equal to marginal cost. Finally, each user receives a fraction of the allocated rate in proportion to their bid. It is shown in [65] that if users do not anticipate the effect of their bid on the price, at an equilibrium the resulting allocation maximizes aggregate surplus (i.e., the sum of users' utilities minus the cost of the total allocated rate).

The pricing mechanism of [65] takes as input the bids of the users, and produces as output the price of the link, and the resulting rate allocation to the users. Kelly et al. continue on to discuss distributed algorithms for implementation of this market-clearing process [65]: given the bids of the users, Kelly et al. present two algorithms which converge to the market-clearing price and rate allocation. Indeed, much of the interest in this market mechanism stems from its desirable properties as a decentralized system, including both stability and scalability. For details, we refer the reader to [59, 64, 127, 140].

One important interpretation of the price signal given to users in the algorithms of [65] is that it can be used to provide early notification of congestion at links in the network. Building on the Explicit Congestion Notification (ECN) proposal [102], this interpretation suggests that the network might charge users proactively, in hopes of avoiding congestion at links later. From an implementation standpoint, such a shift implies that rather than a hard capacity constraint (i.e., a link is overloaded when the rate through it exceeds the capacity of the link), the link has an elastic capacity (i.e., the link gradually begins to signal a buildup of congestion before the link's true capacity is actually met). Many proposals for "active queue management" (AQM) to achieve good performance with Explicit Congestion Notification have been made; see, e.g., [6, 63, 71, 72]. This issue is of secondary importance to our discussion, as we do not concern ourselves with the specific interpretation of the cost function at the link (though an insightful discussion of the relationship between active queue management and the cost function of the link may be found in [45]).

In this chapter, we investigate the robustness of the market mechanism of [65] when users attempt to manipulate the market. Formally, we consider a model where users anticipate the effects of their actions on link prices. This makes the model a game, and we ask two fundamental questions: first, does a Nash equilibrium exist for this game? And second, how inefficient is such an equilibrium relative to the maximal aggregate surplus? We show that Nash equilibria exist, and that the efficiency loss is no more than approximately 34% when users are price anticipating.

In addition to considering the "proportional" allocation mechanism of the previous chapter, the elasticity of supply allows us to consider another well-known model for resource allocation: *Cournot competition*. In Cournot competition [23], the strategy of each user is the quantity of rate that they desire, rather than the total payment they are willing to make (as in the scheme of [65]). Cournot competition is not well-defined when the supply is inelastic, since of course no market-clearing price will exist if the aggregate demanded rate of the users exceeds the supply available at the resource. On the other hand, Cournot games are one of the most well studied economic models for competition among market participants. Historically the focus on Cournot competition has been on Cournot *oligopoly*, i.e., the competition between multiple firms to satisfy an elastic demand—indeed, this was the original model studied by Cournot in 1838 [23]. (For surveys of this rich topic, see [29, 41, 119].) By contrast, in this chapter we consider Cournot *oligopsony*, i.e., the competition between multiple consumers for a single resource in elastic supply. Such models have been previously considered in the context of labor markets, where a small number of firms compete for an available supply of workers [80]. We consider Cournot competition as a model for allocation of rate at a link, in contrast to the proportional allocation of [62, 65]. We will show that in general, the efficiency loss is arbitrarily high under Cournot competition. However, in certain special cases, we will find that the efficiency loss is guaranteed to be no larger than 33%.

Chapter Outline

The remainder of the chapter is organized as follows. We start by considering the pricing mechanism of [65] for a single link. In Section 3.1, we describe the market mechanism for a single link, and recapitulate the results of Kelly et al. [65]. In Section 3.1.2, we describe a game where users of a single link are price anticipating, and establish the existence of a Nash equilibrium. We also establish necessary and sufficient conditions for a strategy vector to be a Nash equilibrium, and in Section 3.1.3 we prove, under an additional assumption on the cost function of the link, that the Nash equilibrium is unique. In Section 3.2 we prove the main result of the chapter for a single link: that when users are price anticipating, the efficiency loss—that is, the loss in aggregate surplus relative to the maximum—is no more than 34%.

In Section 3.3, we compare the settings of inelastic and elastic supply. In particular, we consider a limit of cost functions which approach a hard capacity constraint. We show that if these cost functions are monomials and we let the exponent approach infinity, then the worst case efficiency loss approaches 25%, which is consistent with the result of the previous chapter (Theorem 2.6).

In Section 3.4, we extend the results to general networks. This extension is achieved using the same approach as Section 2.4. We consider a game where users submit individual bids to each link in the network, and establish existence of a Nash equilibrium. We also show, using methods similar to the proof of Theorem 2.13, that the efficiency loss is no more than 34% when users are price anticipating (matching the result of Section 3.2).

In Section 3.5, we turn our attention to Cournot competition. We first present the basic model of Cournot oligopsony, where multiple consumers of a resource in elastic supply choose the quantity they wish to consume. The price of the resource is then set equal to the marginal cost of the total requested allocation. We show that in general, the efficiency loss of such a scheme can be arbitrarily high when users are price anticipating. However, in Section 3.5.1 we consider several special cases and show that efficiency loss is no more than $1/3$ in each of these cases. We show that if R users with the same utility function compete for a resource with a differentiable marginal cost function, then the efficiency loss is no more than $1/(2R+1)$ when the users are price anticipating; we also establish that if the marginal cost function is not differentiable, the efficiency loss is no more than $1/3$ if exactly one user is bidding for the resource. If users may have arbitrary utility functions, we show that the efficiency loss is no more than $1/3$ if the marginal cost function is linear (i.e., the cost function is quadratic). We consider this to be the most applicable of the results of this section, since users will generally not have identical utility functions. We also show in Section 3.5.2 that such a result extends to a network setting, with the additional appealing feature that users only need to choose rates on paths through the network, rather than bidding to individual links as in the mechanism of Section 3.4.

In Section 3.6, we consider the possibility that utility may depend on more than just the rate allocated to a user of a communication network; the utility to a user may be decreased if his traffic experiences a high *latency* passing through the network. To investigate this phenomenon, we extend the Cournot competition model to include the possibility that users' utilities depend on both rate and latency. We show that, even when users act as price takers and latency takers, the efficiency loss may be arbitrarily high if the link manager is restricted to setting a price only as a function of the total rate allocated at the link (an assumption satisfied by the pricing models of both Section 3.1 and 3.5). We then note that this result gives insight into the efficiency loss in the "selfish routing" model of Wardrop [144] when the demand of users is elastic. This example is a complement to the recent results of Roughgarden and Tardos [107, 108], who consider loss of efficiency in selfish routing models when the demand of users is inelastic.

■ 3.1 Preliminaries

Suppose R users share a single communication link. Let $d_r \geq 0$ denote the rate allocated to user r. We assume that user r receives a *utility* equal to $U_r(d_r)$ if the allocated rate is d_r. In addition, we let $f = \sum_r d_r$ denote the total rate allocated at the link, and let $C(f)$ denote the cost incurred at the link when the total allocated rate is $f \geq 0$. We will assume that both U_r and C are measured in the same monetary units. A natural interpretation is that $U_r(d_r)$ is the monetary value to user r of a rate allocation d_r, and $C(f)$ is a monetary cost for congestion at the link when the total allocated rate is f.

We make the following assumptions regarding U_r and C; note that Assumption 3.1 is identical to Assumption 2.1.

Assumption 3.1
For each r, over the domain $d_r \geq 0$ the utility function $U_r(d_r)$ is concave, strictly increasing, and continuous; and over the domain $d_r > 0$, $U_r(d_r)$ is continuously differentiable. Furthermore, the right directional derivative at 0, denoted $U_r'(0)$, is finite.

Assumption 3.2
There exists a continuous, convex, strictly increasing function $p(f)$ over $f \geq 0$ with $p(0) = 0$, such that for $f \geq 0$:

$$C(f) = \int_0^f p(z)dz.$$

Thus $C(f)$ is strictly convex and strictly increasing.

We note that we make rather strong differentiability assumptions here on the utility functions; these assumptions are primarily made to ease the presentation. In Section 3.4, we will relax the assumption that U_r is differentiable.

The condition that $p(0) = 0$ in Assumption 3.2 plays an important role in the subsequent development. We relax this assumption in the context of Cournot competition (Section 3.5). We also present there an example that shows efficiency loss may be arbitrarily high if users are price anticipating under the mechanism of this section, and $p(0) > 0$; see Example 3.3.

Given complete knowledge and centralized control of the system, a natural problem for the network manager to try to solve is the following [62]:

SYSTEM:

$$\text{maximize} \quad \sum_r U_r(d_r) - C\left(\sum_r d_r\right) \tag{3.1}$$

$$\text{subject to} \quad d_r \geq 0, \quad r = 1, \ldots, R. \tag{3.2}$$

We refer to the objective function (3.1) as the *aggregate surplus*; see Section 1.1. This is the net monetary benefit to the economy consisting of the users and the single link. Since the objective function is continuous, and U_r increases at most linearly while C increases superlinearly, an optimal solution $\mathbf{d}^S = (d_1^S, \ldots, d_R^S)$ exists; since the feasible region is convex and C is strictly convex, if the functions U_r are strictly concave, then the optimal solution is unique.

In general, the utility functions are not available to the resource manager. As a result, we consider the following pricing scheme for rate allocation. Each user r makes a payment (also called a *bid*) of w_r to the resource manager. Given the composite vector $\mathbf{w} = (w_1, \ldots, w_r)$, the resource manager chooses a rate allocation $\mathbf{d}(\mathbf{w}) = (d_1(\mathbf{w}), \ldots, d_R(\mathbf{w}))$. We assume the manager treats all users alike—in other words, the network manager does not *price differentiate*. Thus the network manager sets a single price $\mu(\mathbf{w})$; we assume that $\mu(\mathbf{w}) = 0$ if $w_r = 0$ for all r, and $\mu(\mathbf{w}) > 0$ otherwise. All users are then charged the same price $\mu(\mathbf{w})$, leading to:

$$d_r(\mathbf{w}) = \begin{cases} 0, & \text{if } w_r = 0; \\ \dfrac{w_r}{\mu(\mathbf{w})}, & \text{if } w_r > 0. \end{cases}$$

Notice that, with this formulation, the rate allocated to user r to user r is similar to the rate allocated to user r in the model of Section 2.1. The key difference in this setting is that the aggregate rate is not constrained to an inelastic supply; rather, associated with the choice of price $\mu(\mathbf{w})$ is an aggregate rate function $f(\mathbf{w})$, defined by:

$$f(\mathbf{w}) = \sum_r d_r(\mathbf{w}) = \begin{cases} 0, & \text{if } \sum_r w_r = 0; \\ \dfrac{\sum_r w_r}{\mu(\mathbf{w})}, & \text{if } \sum_r w_r > 0. \end{cases} \tag{3.3}$$

We will assume that w_r is measured in the same monetary units as both U_r and C. In this case, given a price $\mu > 0$, user r wishes to maximize the following payoff function over $w_r \geq 0$:

$$P_r(w_r; \mu) = U_r\left(\frac{w_r}{\mu}\right) - w_r. \tag{3.4}$$

The first term represents the utility to user r of receiving a rate allocation equal to w_r/μ; the second term is the payment w_r made to the manager.

Notice that as formulated above, the payoff function P_r assumes that user r acts as a *price taker*; that is, user r does not *anticipate* the effect of his choice of w_r on the price μ, and hence on his resulting rate allocation $d_r(\mathbf{w})$. Informally, we expect that in such a situation the aggregate surplus will be maximized if the network manager sets a price equal to marginal cost; that is, if the price function satisfies:

$$\mu(\mathbf{w}) = p(f(\mathbf{w})). \tag{3.5}$$

We show in the following proposition that a joint solution to (3.3) and (3.5) can be found; we then use this proposition to show that when users optimize (3.4) and the price is set to satisfy (3.5), aggregate surplus is maximized.

Proposition 3.1
Suppose Assumption 3.2 holds. Given any vector of bids $\mathbf{w} \geq 0$, there exists a unique pair $(\mu(\mathbf{w}), f(\mathbf{w})) \geq 0$ satisfying (3.3) *and* (3.5), *and in this case $f(\mathbf{w})$ is the unique solution f to:*

$$\sum_r w_r = fp(f). \tag{3.6}$$

Furthermore, $f(\cdot)$ has the following properties: (1) $f(\mathbf{0}) = 0$; (2) $f(\mathbf{w})$ is continuous for $\mathbf{w} \geq 0$; (3) $f(\mathbf{w})$ is a strictly increasing and strictly concave function of $\sum_r w_r$; and (4) $f(\mathbf{w}) \to \infty$ as $\sum_r w_r \to \infty$.

Proof. Fix a vector $\mathbf{w} \geq 0$. First suppose there exists a solution to (3.3) and (3.5). Then from (3.3), we have:

$$\sum_r w_r = f(\mathbf{w})\mu(\mathbf{w}).$$

After substituting (3.5), this becomes the equation (3.6). Conversely, if $f(\mathbf{w})$ solves (3.6), then defining $\mu(\mathbf{w})$ according to (3.5) makes (3.6) equivalent to (3.3).

Thus, it suffices to check that there exists a unique solution f to (3.6). By Assumption 3.2, p is strictly increasing, and since p is convex, $p(f) \to \infty$ as $f \to \infty$; thus defining $g(f) = fp(f)$, we know $g(0) = 0$; g is strictly increasing, strictly convex, and continuous; and $g(f) \to \infty$ as $f \to \infty$. Thus g is invertible, and crosses the level $\sum_r w_r$ at a unique value $f(\mathbf{w}) = g^{-1}(\sum_r w_r)$. From this description and the properties of g we immediately see that f has the four properties described in the proposition. □

Observe that we can view (3.6) as a market-clearing process. Given the total revenue $\sum_r w_r$ from the users, the link manager chooses an aggregate rate $f(\mathbf{w})$ so that the revenue is exactly equal to the aggregate charge $f(\mathbf{w})p(f(\mathbf{w}))$. Due to Assumption 3.2, this market-clearing aggregate rate is uniquely determined. Kelly et al. present two algorithms in [65] which amount to dynamic processes of market-clearing; as a result, a key motivation for the mechanism we study in this chapter is that it represents the equilibrium behavior of the algorithms in [65].

Indeed, as in Section 2.1, we can view this market-clearing process in terms of supply and demand. To see this interpretation, note that as in Section 2.1, when a user chooses a total payment w_r, it is as if the user has chosen a *demand function* $D(\overline{p}, w_r) = w_r/\overline{p}$ for $\overline{p} > 0$. The demand function describes the amount of rate the user demands at any given price $\overline{p} > 0$. In contrast to Chapter 2, however, the available supply of rate at the link is *elastic*. In particular, we define the *supply function* $S(\overline{p})$ of the link manager as the inverse of the price function p: $S(\overline{p}) = p^{-1}(\overline{p})$. The link manager then chooses a price $\mu > 0$ so that $\sum_r D(\mu, w_r) = S(\mu)$, i.e., so that the aggregate demand equals the available supply $S(\mu)$; see Figure 3-1. For the specific form of demand functions we consider here, this leads to the equation for $f(\mathbf{w})$ and $\mu = p(f(\mathbf{w}))$ given in (3.6). Note that we have $S(p(f(\mathbf{w}))) = f(\mathbf{w})$, by definition. User r then receives a rate allocation given by $D(p(f(\mathbf{w})), w_r)$, and makes a payment $p(f(\mathbf{w}))D(p(f(\mathbf{w})), w_r) = w_r$.

In the remainder of the section, we consider two different models for how users might interact with this price mechanism. In Section 3.1.1, we consider a model where users do not anticipate the effect of their bids on the price, and establish existence of a competitive equilibrium (a result due to Kelly et al. [65]). Furthermore, this competitive equilibrium leads to an allocation which is an optimal solution to *SYSTEM*. In Section 3.1.2, we change the model and assume users are price anticipating, and establish existence of a Nash equilibrium. In Section 3.1.3, we show for a certain class of price functions that the Nash equilibrium is in fact unique. Finally, Section 3.2 considers the loss of efficiency at Nash equilibria, relative to the optimal solution to *SYSTEM*.

■ 3.1.1 Price Taking Users and Competitive Equilibrium

Kelly et al. show in [65] that when users are price takers, and the network sets the price $\mu(\mathbf{w})$ according to (3.3) and (3.5), the resulting allocation is an optimal solution to *SYSTEM*. This is formalized in the following theorem, adapted from [65].

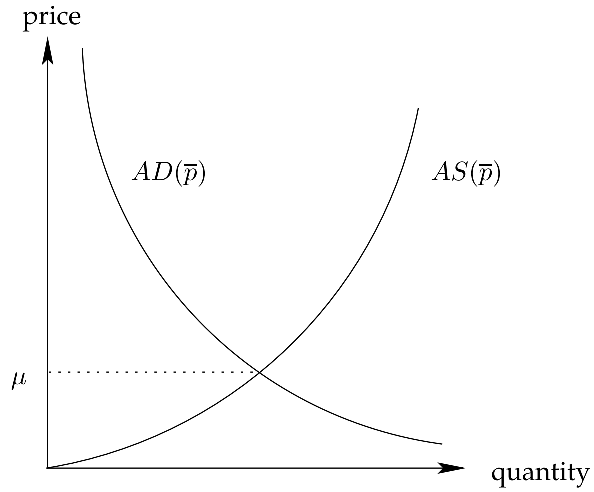

Figure 3-1. The market-clearing process with elastic supply: Each consumer r chooses a willingness-to-pay w_r, which maps to a demand function $D(\overline{p}, w_r) = w_r/\overline{p}$. This defines the aggregate demand function $AD(\overline{p}) = \sum_r D(\overline{p}, w_r) = \sum_r w_r/\overline{p}$. The aggregate supply function is $AS(\overline{p}) = p^{-1}(\overline{p})$. The price μ is chosen so that supply equals demand, i.e., so that $\sum_r w_r/\mu = AD(\mu) = AS(\mu) = p^{-1}(\mu)$.

Theorem 3.2 (Kelly et al., [65])
Suppose Assumptions 3.1 and 3.2 hold. Given $\mathbf{w} \geq 0$, *let* $(\mu(\mathbf{w}), f(\mathbf{w}))$ *be the unique solution to* (3.3) *and* (3.5). *Then there exists a vector* $\mathbf{w}$ *such that* $\mu(\mathbf{w}) > 0$, *and:*

$$P_r(w_r; \mu(\mathbf{w})) = \max_{\overline{w}_r \geq 0} P_r(\overline{w}_r; \mu(\mathbf{w})), \quad r = 1, \ldots, R. \tag{3.7}$$

For any such vector $\mathbf{w}$, *the vector* $\mathbf{d}(\mathbf{w}) = \mathbf{w}/\mu(\mathbf{w})$ *is an optimal solution to SYSTEM. If the functions* U_r *are strictly concave, such a vector* $\mathbf{w}$ *is unique as well.*

Proof. Let $\mathbf{d}^S$ be any optimal solution to *SYSTEM*; as discussed above, at least one such solution exists. Let $f^S = \sum_r d_r^S$, and define $w_r^S = d_r^S p(f^S)$ for each r. Observe that with this definition, we have $\sum_r w_r^S = \sum_r d_r^S p(f^S) = f^S p(f^S)$; thus f^S satisfies (3.6), and we have $f(\mathbf{w}^S) = f^S$, $\mathbf{d}(\mathbf{w}^S) = \mathbf{d}^S$.

Given Assumptions 3.1 and 3.2, observe that any optimal solution to *SYSTEM* is identified by the following necessary and sufficient optimality conditions:

$$U_r'(d_r^S) = p\left(\sum_s d_s^S\right), \quad \text{if } d_r^S > 0; \tag{3.8}$$

$$U_r'(0) \leq p\left(\sum_s d_s^S\right), \quad \text{if } d_r^S = 0. \tag{3.9}$$

Now since $p(0) = 0$ but $U_r'(0) > 0$ for all r, we must have $f^S = \sum_r d_r^S > 0$; thus $\mu(\mathbf{w}) = p(f^S) > 0$. But then $d_r^S = w_r/p(f^S)$ for each r, so the preceding optimality conditions become:

$$U_r'\left(\frac{w_r}{p(f^S)}\right) = p(f^S), \quad \text{if } w_r > 0;$$
$$U_r'(0) \le p(f^S), \quad \text{if } w_r = 0.$$

These conditions ensure that (3.7) holds.

Conversely, suppose we are given a vector $\mathbf{w}$ such that $\mu(\mathbf{w}) > 0$, and (3.7) holds. Then we simply reverse the argument above: since (3.7) holds, we conclude that the optimality conditions (3.8)-(3.9) hold with $\mathbf{d}(\mathbf{w}) = \mathbf{w}/\mu(\mathbf{w}) = \mathbf{w}/p(f(\mathbf{w}))$, so that $\mathbf{d}(\mathbf{w})$ is an optimal solution to *SYSTEM*. Finally, if the functions U_r are each strictly concave, then the optimal solution $\mathbf{d}^S$ to *SYSTEM* is unique, so the price $p(f^S)$ is uniquely determined as well. As a result, for each r the product $d_r^S p(f^S)$ is unique, so the vector $\mathbf{w}$ identified in the theorem must be unique as well. □

Theorem 3.2 shows that with an appropriate choice of price function (as determined by (3.3) and (3.5)), and under the assumption that the users of the link behave as price takers, there exists a bid vector $\mathbf{w}$ where all users have optimally chosen their bids w_r, with respect to the given price $\mu(\mathbf{w})$; and at this "equilibrium," the aggregate surplus is maximized. However, when the price taking assumption is violated, the model changes into a game and the guarantee of Theorem 3.2 is no longer valid. We investigate this game in the following section.

■ 3.1.2 Price Anticipating Users and Nash Equilibrium

We now consider an alternative model where the users of a single link are price anticipating, rather than price taking, and play a game to acquire a share of the link. Throughout the remainder of this section as well as in Section 3.2, we will assume that the link manager sets the price $\mu(\mathbf{w})$ according to the unique choice prescribed by Proposition 3.1, as follows.

Assumption 3.3
For all $\mathbf{w} \ge 0$, the aggregate rate $f(\mathbf{w})$ is the unique solution to (3.6)*: $\sum_r w_r = f(\mathbf{w})p(f(\mathbf{w}))$. Furthermore, for each r, $d_r(\mathbf{w})$ is given by:*

$$d_r(\mathbf{w}) = \begin{cases} 0, & \text{if } w_r = 0; \\ \dfrac{w_r}{p(f(\mathbf{w}))}, & \text{if } w_r > 0. \end{cases} \tag{3.10}$$

Note that we have $f(\mathbf{w}) > 0$ and $p(f(\mathbf{w})) > 0$ if $\sum_r w_r > 0$, and hence d_r is always well defined.

We adopt the notation $\mathbf{w}_{-r}$ to denote the vector of all bids by users other than r; i.e., $\mathbf{w}_{-r} = (w_1, w_2, \ldots, w_{r-1}, w_{r+1}, \ldots, w_R)$. Then given $\mathbf{w}_{-r}$, each user r chooses $w_r \geq 0$ to maximize:

$$Q_r(w_r; \mathbf{w}_{-r}) = U_r(d_r(\mathbf{w})) - w_r, \tag{3.11}$$

over nonnegative w_r. The payoff function Q_r is similar to the payoff function P_r, except that the user now anticipates that the network will set the price according to Assumption 3.3, as captured by the allocated rate $d_r(\mathbf{w})$. A *Nash equilibrium* of the game defined by $(Q_1, \ldots, Q_R)$ is a vector $\mathbf{w} \geq 0$ such that for all r:

$$Q_r(w_r; \mathbf{w}_{-r}) \geq Q_r(\overline{w}_r; \mathbf{w}_{-r}), \quad \text{for all } \overline{w}_r \geq 0. \tag{3.12}$$

We begin by asking whether a Nash equilibrium exists for the game defined by $(Q_1, \ldots, Q_R)$. We will need the following proposition.

Proposition 3.3
Suppose that Assumptions 3.1-3.3 hold. Then: (1) $d_r(\mathbf{w})$ is a continuous function of $\mathbf{w}$; and (2) for any $\mathbf{w}_{-r} \geq 0$, $d_r(\mathbf{w})$ is strictly increasing and concave in $w_r \geq 0$, and $d_r(\mathbf{w}) \to \infty$ as $w_r \to \infty$.

Proof. We first show that $d_r(\mathbf{w})$ is a continuous function of $\mathbf{w}$. Recall from Proposition 3.1 that $f(\mathbf{w})$ is a continuous function of $\mathbf{w}$, and $f(\mathbf{0}) = 0$. Now at any vector $\mathbf{w}$ such that $\sum_s w_s > 0$, we have $p(f(\mathbf{w})) > 0$, so $d_r(\mathbf{w}) = w_r/p(f(\mathbf{w}))$; thus continuity of d_r at $\mathbf{w}$ follows by continuity of f and p. Suppose instead that $\mathbf{w} = \mathbf{0}$, and consider a sequence $\mathbf{w}(n) \to \mathbf{0}$ as $n \to \infty$. Then $\sum_r d_r(\mathbf{w}(n)) = f(\mathbf{w}(n)) \to 0$ as $n \to \infty$, from parts (1) and (2) of Proposition 3.1; since $d_r(\mathbf{w}(n)) \geq 0$ for all n, we must have $d_r(\mathbf{w}(n)) \to 0 = d_r(\mathbf{0})$ as $n \to \infty$, as required.

From Assumption 3.3, we can rewrite the definition of $d_r(\mathbf{w})$ as:

$$d_r(\mathbf{w}) = \begin{cases} 0, & \text{if } w_r = 0; \\ \dfrac{w_r}{\sum_s w_s} f(\mathbf{w}), & \text{if } w_r > 0. \end{cases} \tag{3.13}$$

From this expression and Proposition 3.1, it follows that $d_r(\mathbf{w})$ is a continuous function of $\mathbf{w}$, and that $d_r(\mathbf{w})$ is strictly increasing in w_r. To show $d_r(\mathbf{w}) \to \infty$ as $w_r \to \infty$, we only need $f(\mathbf{w}) \to \infty$ as $w_r \to \infty$, a fact that was shown in Proposition 3.1.

It remains to be shown that for fixed $\mathbf{w}_{-r}$, d_r is a concave function of $w_r \geq 0$. Since we have already shown that d_r is continuous, we may assume without loss of generality that $w_r > 0$. We first assume that p is twice differentiable. In this case, it follows from (3.6) that f is twice differentiable in w_r. Since $w_r > 0$, we can differentiate

(3.13) twice to find:

$$\frac{\partial^2 d_r(\mathbf{w})}{\partial w_r^2} = -\left(\frac{2\sum_{s\neq r} w_s}{(\sum_s w_s)^3}\right) f(\mathbf{w}) + \left(\frac{2\sum_{s\neq r} w_s}{(\sum_s w_s)^2}\right) \frac{\partial f(\mathbf{w})}{\partial w_r} + \left(\frac{w_r}{\sum_s w_s}\right) \frac{\partial^2 f(\mathbf{w})}{\partial w_r^2}.$$

From Proposition 3.1, f is a strictly concave function of $\sum_s w_s$; thus the last term in the sum above is nonpositive. To show that d_r is concave in w_r, therefore, it suffices to show that the sum of the first two terms is negative, i.e.:

$$\frac{f(\mathbf{w})}{\sum_s w_s} \geq \frac{\partial f(\mathbf{w})}{\partial w_r}.$$

By differentiating both sides of (3.6), we find that:

$$\frac{\partial f(\mathbf{w})}{\partial w_r} = \frac{1}{p(f(\mathbf{w})) + f(\mathbf{w})p'(f(\mathbf{w}))}.$$

On the other hand, from (3.6), we have:

$$\frac{f(\mathbf{w})}{\sum_s w_s} = \frac{1}{p(f(\mathbf{w}))}.$$

Substituting these relations, and noting that $f(\mathbf{w})p'(f(\mathbf{w})) \geq 0$ since p is strictly increasing, we have:

$$\frac{f(\mathbf{w})}{\sum_s w_s} = \frac{1}{p(f(\mathbf{w}))} \geq \frac{1}{p(f(\mathbf{w})) + f(\mathbf{w})p'(f(\mathbf{w}))} = \frac{\partial f(\mathbf{w})}{\partial w_r},$$

as required. Thus $d_r(\mathbf{w})$ is concave in w_r, as long as p is twice differentiable.

Now suppose that p is any price function satisfying Assumption 3.2, but not necessarily twice differentiable. In this case, we choose a sequence of twice differentiable price functions p^n satisfying Assumption 3.2, such that $p^n \to p$ pointwise as $n \to \infty$ (i.e., $p^n(f) \to p(f)$ as $n \to \infty$, for all $f \geq 0$). We define $p^n(f)$ as follows. Define $p(f) = 0$ for $f \leq 0$, and consider a sequence of twice differentiable functions ϕ^n, such that ϕ^n has support on $[-1/n, 0]$, and $\int_{-\infty}^{\infty} \phi^n(z)\ dz = 1$. Then it is straightforward to verify the sequence p^n defined by $p^n(f) = \int_{-\infty}^{\infty} p(z)\phi^n(z - f)\ dz$ has the required properties.

Let d_r^n be the allocation function for user r when the price function is p^n; then $d_r^n(\mathbf{w})$ is concave in w_r, for each n. In order to show that $d_r(\mathbf{w})$ is concave in w_r, therefore, it suffices to show that $d_r^n \to d_r$ pointwise as $n \to \infty$. From (3.13), this will be true as long as $f^n \to f$ pointwise as $n \to \infty$, where f^n is the solution to (3.6) when the price function is p^n.

Fix a bid vector $\mathbf{w}$; we now proceed to show that $f^n(\mathbf{w}) \to f(\mathbf{w})$ as $n \to \infty$. For each n, define $g^n(f) = fp^n(f)$, and let $g(f) = fp(f)$. Then $g^n(f) \to g(f)$ as $n \to \infty$, for all $f \geq 0$. Furthermore, from (3.6), $\sum_r w_r = g(f(\mathbf{w}))$. Fix $\varepsilon > 0$, and choose $\delta > 0$ so that:

$$\delta < \min\{\sum_r w_r - g(f(\mathbf{w}) - \varepsilon), g(f(\mathbf{w}) + \varepsilon) - \sum_r w_r\}.$$

(Note that such a choice is possible because g is strictly increasing.) Now for sufficiently large n, we will have:

$$g^n(f(\mathbf{w}) - \varepsilon) - g(f(\mathbf{w}) - \varepsilon) < \delta, \quad g(f(\mathbf{w}) + \varepsilon) - g^n(f(\mathbf{w}) + \varepsilon) < \delta.$$

From the definition of δ, this yields:

$$g^n(f(\mathbf{w}) - \varepsilon) < \sum_r w_r < g^n(f(\mathbf{w}) + \varepsilon).$$

Since g^n is strictly increasing, and $f^n(\mathbf{w})$ satisfies $g^n(f^n(\mathbf{w})) = \sum_r w_r$, we conclude that $|f^n(\mathbf{w}) - f(\mathbf{w})| < \varepsilon$ for sufficiently large n, as required. $\square$

The previous proposition establishes concavity and continuity of d_r; this guarantees existence of a Nash equilibrium, as the following proposition shows.

Proposition 3.4

Suppose that Assumptions 3.1-3.3 hold. Then there exists a Nash equilibrium $\mathbf{w}$ *for the game defined by* $(Q_1, \ldots, Q_R)$*.*

Proof. We begin by observing that we may restrict the strategy space of each user r to a compact set, without loss of generality. To see this, fix a user r, and a vector $\mathbf{w}_{-r}$ of bids for all other users. Given a bid w_r for user r, we note that $d_r(\mathbf{w}) \leq d_r(w_r; \mathbf{0}_{-r})$, where $\mathbf{0}_{-r}$ denotes the bid vector where all other users bid zero. This inequality follows since $w_r = d_r(\mathbf{w})p(f(\mathbf{w}))$; and if $\sum_s w_s$ decreases, then $p(f(\mathbf{w}))$ decreases as well (from Proposition 3.1), so $d_r(\mathbf{w})$ must increase.

We thus have $Q_r(w_r; \mathbf{w}_{-r}) \leq U_r(d_r(w_r; \mathbf{0}_{-r})) - w_r$. By concavity of U_r, for $w_r > 0$ we have:

$$Q_r(w_r; \mathbf{w}_{-r}) \leq U_r(0) + U_r'(0)d_r(w_r; \mathbf{0}_{-r}) - w_r = U_r(0) + w_r\left(\frac{U_r'(0)}{p(f(w_r; \mathbf{0}_{-r}))} - 1\right). \tag{3.14}$$

Now observe from (3.6) that:

$$w_r = f(w_r; \mathbf{0}_{-r})p(f(w_r; \mathbf{0}_{-r})).$$

Since p is convex and strictly increasing, we have $\lim_{f\to\infty} p(f) = \infty$; thus we conclude that $p(f(w_r; \mathbf{0}_{-r})) \to \infty$ as $w_r \to \infty$. Consequently, using (3.14), there exists $B_r > 0$ such that if $w_r \geq B_r$, then $Q_r(w_r; \mathbf{w}_{-r}) < U_r(0)$. Since $Q_r(0; \mathbf{w}_{-r}) = U_r(0)$, user r

would never choose to bid $w_r \geq B_r$ at a Nash equilibrium. Thus, we may restrict the strategy space of user r to the compact interval $S_r = [0, B_r]$ without loss of generality.

The game defined by $(Q_1, \ldots, Q_R)$ together with the strategy spaces $(S_1, \ldots, S_R)$ is now a *concave R-person game*: applying Proposition 3.3, each payoff function Q_r is continuous in the composite strategy vector $\mathbf{w}$, and concave in w_r (since U_r is concave and strictly increasing, and $d_r(\mathbf{w})$ is concave in w_r); and the strategy space of each user r is a compact, convex, nonempty subset of $\mathbb{R}$. Applying Rosen's existence theorem [104] (proven using Kakutani's fixed point theorem), we conclude that a Nash equilibrium $\mathbf{w}$ exists for this game. □

In the remainder of this section, we will establish necessary and sufficient conditions for a vector $\mathbf{w}$ to be a Nash equilibrium. Because the price function p may not be differentiable, we will require elements of the theory of *subgradients* to describe necessary local conditions for a vector $\mathbf{w}$ to be a Nash equilibrium. Since the payoff of user r is concave, these necessary conditions will in fact be sufficient for $\mathbf{w}$ to be a Nash equilibrium. (The reader is referred to the Notation section for necessary details on subgradients.)

For the remainder of the chapter, we view any price function p as an extended real-valued convex function, by defining $p(f) = \infty$ for $f < 0$. Our first step is a lemma demonstrating that d_r is directionally differentiable (as a function of w_r); for notational convenience, we will require the following definitions of $\varepsilon^+(f)$ and $\varepsilon^-(f)$, for $f > 0$:

$$\varepsilon^+(f) = \frac{f}{p(f)} \cdot \frac{\partial^+ p(f)}{\partial f}, \qquad \varepsilon^-(f) = \frac{f}{p(f)} \cdot \frac{\partial^- p(f)}{\partial f}. \tag{3.15}$$

Note that under Assumption 3.2, we have $0 < \varepsilon^-(f) \leq \varepsilon^+(f)$ for $f > 0$.

Lemma 3.5 *Suppose Assumptions 3.1-3.3 hold. Then for all $\mathbf{w}$ with $\sum_s w_s > 0$, $d_r(\mathbf{w})$ is directionally differentiable with respect to w_r. These directional derivatives are given by:*

$$\frac{\partial^+ d_r(\mathbf{w})}{\partial w_r} = \frac{1}{p(f(\mathbf{w}))}\left(1 - \frac{d_r(\mathbf{w})}{f(\mathbf{w})} \cdot \frac{\varepsilon^+(f(\mathbf{w}))}{1 + \varepsilon^+(f(\mathbf{w}))}\right); \tag{3.16}$$

$$\frac{\partial^- d_r(\mathbf{w})}{\partial w_r} = \frac{1}{p(f(\mathbf{w}))}\left(1 - \frac{d_r(\mathbf{w})}{f(\mathbf{w})} \cdot \frac{\varepsilon^-(f(\mathbf{w}))}{1 + \varepsilon^-(f(\mathbf{w}))}\right). \tag{3.17}$$

Furthermore, $\partial^+ d_r(\mathbf{w})/\partial w_r > 0$, and if $w_r > 0$ then $\partial^- d_r(\mathbf{w})/\partial w_r > 0$.

Proof. Existence of the directional derivatives is obtained because $d_r(\mathbf{w})$ is a concave function of w_r (from Proposition 3.3). Fix a vector $\mathbf{w}$ of bids, such that $\sum_r w_r > 0$. Since f is an increasing, concave function of w_r, and the convex function p is directionally differentiable at $f(\mathbf{w})$ ([103], Theorem 23.1), we can apply the chain rule to compute the right directional derivative of (3.6) with respect to w_r:

$$1 = \frac{\partial^+ f(\mathbf{w})}{\partial w_r} \cdot p(f(\mathbf{w})) + f(\mathbf{w}) \cdot \frac{\partial^+ p(f(\mathbf{w}))}{\partial f} \cdot \frac{\partial^+ f(\mathbf{w})}{\partial w_r}.$$

Thus, as long as $\sum_r w_r > 0$, $\partial^+ f(\mathbf{w})/\partial w_r$ exists, and is given by:

$$\frac{\partial^+ f(\mathbf{w})}{\partial w_r} = \left(p(f(\mathbf{w})) + f(\mathbf{w}) \cdot \frac{\partial^+ p(f(\mathbf{w}))}{\partial f} \right)^{-1}.$$

We conclude from (3.10) that the right directional derivative of $d_r(\mathbf{w})$ with respect to w_r is given by:

$$\frac{\partial^+ d_r(\mathbf{w})}{\partial w_r} = \frac{1}{p(f(\mathbf{w}))} - \frac{w_r}{\big(p(f(\mathbf{w}))\big)^2} \cdot \frac{\partial^+ p(f(\mathbf{w}))}{\partial f} \cdot \frac{\partial^+ f(\mathbf{w})}{\partial w_r}.$$

Simplifying, this reduces to (3.16). Note that since $d_r(\mathbf{w}) \leq f(\mathbf{w})$ and $\varepsilon^+(f(\mathbf{w}))/[1 + \varepsilon^+(f(\mathbf{w}))] < 1$, we have $\partial^+ d_r(\mathbf{w})/\partial w_r > 0$. A similar analysis follows for the left directional derivative. □

For notational convenience, we make the following definitions for $f > 0$:

$$\beta^+(f) = \frac{\varepsilon^+(f)}{1 + \varepsilon^+(f)}, \qquad \beta^-(f) = \frac{\varepsilon^-(f)}{1 + \varepsilon^-(f)}. \tag{3.18}$$

Under Assumption 3.2, we have $0 < \beta^-(f) \leq \beta^+(f) < 1$ for $f > 0$.

The next proposition is the central result of this section: it establishes local conditions that are necessary and sufficient for a Nash equilibrium.

Proposition 3.6
Suppose that Assumptions 3.1-3.3 hold. Then $\mathbf{w}$ *is a Nash equilibrium of the game defined by* $(Q_1, \ldots, Q_R)$, *if and only if* $\sum_r w_r > 0$, *and with* $\mathbf{d} = \mathbf{d}(\mathbf{w})$, $f = f(\mathbf{w})$, *the following two conditions hold for all* r*:*

$$U_r'(d_r)\left(1 - \beta^+(f) \cdot \frac{d_r}{f}\right) \leq p(f); \tag{3.19}$$

$$U_r'(d_r)\left(1 - \beta^-(f) \cdot \frac{d_r}{f}\right) \geq p(f), \quad \text{if } d_r > 0. \tag{3.20}$$

Conversely, if $\mathbf{d} \geq 0$ *and* $f > 0$ *satisfy* (3.19)-(3.20), *and* $\sum_r d_r = f$, *then the vector* $\mathbf{w} = p(f)\mathbf{d}$ *is a Nash equilibrium with* $\mathbf{d} = \mathbf{d}(\mathbf{w})$ *and* $f = f(\mathbf{w})$.

Proof. We first show that if $\mathbf{w}$ is a Nash equilibrium, then we must have $\sum_r w_r > 0$. Suppose not; then $w_r = 0$ for all r. Fix a user r; for $\overline{w}_r > 0$, we have $d_r(\overline{w}_r; \mathbf{w}_{-r})/\overline{w}_r =$

$f(\overline{w}_r; \mathbf{w}_{-r})/\overline{w}_r = 1/p(f(\overline{w}_r; \mathbf{w}_{-r}))$, which approaches infinity as $\overline{w}_r \to 0$. Thus we have $\partial^+ d_r(\mathbf{w})/\partial w_r = \infty$, which yields:

$$\frac{\partial^+ Q_r(w_r; \mathbf{w}_{-r})}{\partial w_r} = U_r'(0) \cdot \frac{\partial^+ d_r(\mathbf{w})}{\partial w_r} - 1 = \infty.$$

In particular, an infinitesimal increase of w_r will strictly increase the payoff of user r, so $\mathbf{w} = \mathbf{0}$ cannot be a Nash equilibrium. Thus if $\mathbf{w}$ is a Nash equilibrium, then $\sum_r w_r > 0$.

Now let $\mathbf{w}$ be a Nash equilibrium. We established in Lemma 3.5 that d_r is directionally differentiable in w_r for each r, as long as $\sum_s w_s > 0$. Thus, from (3.12), if $\mathbf{w}$ is a Nash equilibrium, then the following two conditions must hold:

$$\begin{aligned} \frac{\partial^+ Q_r(w_r; \mathbf{w}_{-r})}{\partial w_r} &= U_r'(d_r(\mathbf{w})) \cdot \frac{\partial^+ d_r(\mathbf{w})}{\partial w_r} - 1 \leq 0; \\ \frac{\partial^- Q_r(w_r; \mathbf{w}_{-r})}{\partial w_r} &= U_r'(d_r(\mathbf{w})) \cdot \frac{\partial^- d_r(\mathbf{w})}{\partial w_r} - 1 \geq 0, \quad \text{if } w_r > 0. \end{aligned}$$

We may substitute using Lemma 3.5 to find that if $\mathbf{w}$ is a Nash equilibrium, then:

$$\begin{aligned} U_r'(d_r(\mathbf{w})) \left(1 - \beta^+(f(\mathbf{w})) \cdot \frac{d_r(\mathbf{w})}{f(\mathbf{w})}\right) &\leq p(f(\mathbf{w})); \\ U_r'(d_r(\mathbf{w})) \left(1 - \beta^-(f(\mathbf{w})) \cdot \frac{d_r(\mathbf{w})}{f(\mathbf{w})}\right) &\geq p(f(\mathbf{w})), \quad \text{if } w_r > 0. \end{aligned}$$

Since the condition $w_r > 0$ is identical to the condition $d_r(\mathbf{w}) > 0$, this establishes the conditions in the proposition. Conversely, if $\sum_r w_r > 0$ and the preceding two conditions hold, then we may reverse the argument: since the payoff function of user r is a concave function of w_r for each r (from Proposition 3.3), (3.19)-(3.20) are sufficient for $\mathbf{w}$ to be a Nash equilibrium.

Finally, suppose that $\mathbf{d}$ and $f > 0$ satisfy (3.19)-(3.20), with $\sum_r d_r = f$. Then let $w_r = d_r p(f)$. We then have $\sum_r w_r > 0$ (since $f > 0$); and $\sum_r w_r = f p(f)$, so that $f = f(\mathbf{w})$. Finally, since $f > 0$, we have $d_r = w_r/p(f) = w_r/p(f(\mathbf{w}))$, so that $d_r = d_r(\mathbf{w})$. Thus $\mathbf{w}$ is a Nash equilibrium, as required. □

Note that the preceding proposition identifies a Nash equilibrium entirely in terms of the allocation made; and conversely, if we find a pair $(\mathbf{d}, f)$ which satisfies (3.19)-(3.20) with $f > 0$ and $\sum_r d_r = f$, then there exists a Nash equilibrium which yields that allocation. In particular, the set of allocations $\mathbf{d}$ which can arise at Nash equilibria coincides with those vectors $\mathbf{d}$ such that $f = \sum_r d_r > 0$, and (3.19)-(3.20) are satisfied.

3.1.3 Nondecreasing Elasticity Price Functions: Uniqueness of Nash Equilibrium

In this section, we demonstrate that for a certain class of differentiable price functions, there exists a *unique* Nash equilibrium of the game defined by $(Q_1, \ldots, Q_R)$. We consider price functions p which satisfy the following additional assumption.

Assumption 3.4
The price function p *is differentiable, and exhibits* nondecreasing elasticity, *i.e., for* $0 < f_1 \leq f_2$*, there holds:*

$$\frac{f_1 p'(f_1)}{p(f_1)} \leq \frac{f_2 p'(f_2)}{p(f_2)}.$$

To gain some intuition for the concept of nondecreasing elasticity, consider a price function p satisfying Assumption 3.2. The quantity $fp'(f)/p(f)$ is known as the *elasticity* of the price function p [137]. Note that the elasticity of $p(f)$ is the derivative of $\ln p(f)$ with respect to $\ln f$. Intuitively, therefore, elasticity measures the percentage change in $p(f)$ which results from a one percent change in f. From this viewpoint, we see that nondecreasing elasticity is equivalent to the requirement that $\ln p(f)$ be a convex function of $\ln f$. (Note that in general, this is not equivalent to the requirement that p is a convex function of f.)

Nondecreasing elasticity can also be interpreted by considering the price function as the inverse of the *supply function* $S(\mu) = p^{-1}(\mu)$; recall that the supply function gives the amount of rate the link manager is willing to supply at a given price. In this case, nondecreasing elasticity of the price function is equivalent to nonincreasing elasticity of the supply function.

Nondecreasing elasticity captures a wide range of price functions; we give two common examples below.

Example 3.1 (The M/M/1 Queue)
Consider the cost function $C(f) = af/(s - f)$, where $a > 0$ and $s > 0$ are constants; then the cost is proportional to the steady-state queue size in an M/M/1 queue with service rate s and arrival rate f. (Note that we must view p as an extended real-valued function, with $p(f) = \infty$ for $f > s$; this does not affect any of the analysis of this chapter.) It is straightforward to check that, as long as $0 < f < s$ (the region of interest for our purposes), we have:

$$\frac{fp'(f)}{p(f)} = \frac{2f}{s - f},$$

which is a strictly increasing function of f, and approaches ∞ as $f \to s$. Thus p satisfies Assumption 3.4. □

Example 3.2 (M/M/1 Overflow Probability)
Consider the function $p(f) = a(f/s)^B$, where $a > 0$, $s > 0$, and $B \geq 1$ is an integer.

Then the price is set proportional to the probability that an M/M/1 queue exceeds a buffer level B, when the service rate is s and the arrival rate is f. In this case we have $fp'(f)/p(f) = B$, so that p satisfies Assumption 3.4. □

We now prove the key property of differentiable nondecreasing elasticity price functions in the current development: for such functions, there exists a unique Nash equilibrium for the game defined by $(Q_1, \dots, Q_R)$.

Proposition 3.7

Suppose Assumptions 3.1-3.3 hold. If in addition p *is differentiable and exhibits nondecreasing elasticity (Assumption 3.4 holds), then there exists a unique Nash equilibrium for the game defined by* $(Q_1, \dots, Q_R)$.

Proof. We use the expressions (3.19)-(3.20) to show that the Nash equilibrium is unique under Assumption 3.4. Observe that in this case, from (3.18), we may define $\beta(f) = \beta^+(f) = \beta^-(f)$ for $f > 0$, and conclude that $\mathbf{w}$ is a Nash equilibrium if and only if $\sum_s w_s > 0$ and the following optimality conditions hold:

$$U_r'(d_r(\mathbf{w}))\left(1 - \beta(f(\mathbf{w})) \cdot \frac{d_r(\mathbf{w})}{f(\mathbf{w})}\right) = p(f(\mathbf{w})), \quad \text{if } w_r > 0; \tag{3.21}$$

$$U_r'(0) \le p(f(\mathbf{w})), \quad \text{if } w_r = 0. \tag{3.22}$$

Suppose we have two Nash equilibria $\mathbf{w}^1$, $\mathbf{w}^2$, with $0 < \sum_s w_s^1 < \sum_s w_s^2$; then we must have $p(f(\mathbf{w}^1)) < p(f(\mathbf{w}^2))$, and $f(\mathbf{w}^1) < f(\mathbf{w}^2)$. Note that $U_r'(d_r)$ is nonincreasing as d_r increases; and $\beta(f)$ is nondecreasing as f increases (from Assumption 3.4), and therefore $\beta(f(\mathbf{w}^1)) \le \beta(f(\mathbf{w}^2))$. Furthermore, if $w_r^2 > 0$, then from (3.21) we have $U_r'(0) > p(f(\mathbf{w}^2))$; thus $U_r'(0) > p(f(\mathbf{w}^1))$, so $w_r^1 > 0$ as well (from (3.22)).

Now note that the right hand side of (3.21) is strictly larger at $\mathbf{w}^1$ than at $\mathbf{w}^2$; thus the left hand side must be strictly larger at $\mathbf{w}^1$ than at $\mathbf{w}^2$ as well. This is only possible if $d_r(\mathbf{w}^1)/f(\mathbf{w}^1) > d_r(\mathbf{w}^2)/f(\mathbf{w}^2)$ for each user r, since we have shown in the preceding paragraph that $f(\mathbf{w}^1) < f(\mathbf{w}^2)$; $U_r'(d_r)$ is nonincreasing as d_r increases; and $\beta(f(\mathbf{w}^1)) \le \beta(f(\mathbf{w}^2))$. Since $f(\mathbf{w}) = \sum_r d_r(\mathbf{w})$, we have:

$$1 = \sum_{r: w_r^2 > 0} \frac{d_r(\mathbf{w}^2)}{f(\mathbf{w}^2)} < \sum_{r: w_r^2 > 0} \frac{d_r(\mathbf{w}^1)}{f(\mathbf{w}^1)} = 1,$$

which is a contradiction. Thus at the two Nash equilibria, we must have $\sum_s w_s^1 = \sum_s w_s^2$, so we can let $f_0 = f(\mathbf{w}^1) = f(\mathbf{w}^2)$, $p_0 = p(f_0)$, and $\beta_0 = \beta(f_0)$. Then all Nash equilibria $\mathbf{w}$ satisfy:

$$U_r'(d_r(\mathbf{w}))\left(1 - \beta_0 \cdot \frac{d_r(\mathbf{w})}{f_0}\right) = p_0, \quad \text{if } w_r > 0; \tag{3.23}$$

$$U_r'(0) \le p_0, \quad \text{if } w_r = 0. \tag{3.24}$$

But now we observe that the left hand side of (3.23) is strictly decreasing in $d_r(\mathbf{w})$, so given p_0, there exists at most one solution $d_r(\mathbf{w})$ to (3.23). Since $w_r = d_r(\mathbf{w})p_0$, this implies the Nash equilibrium $\mathbf{w}$ must be unique. □

3.2 Efficiency Loss: The Single Link Case

We let $\mathbf{d}^S$ denote an optimal solution to *SYSTEM*, and let $\mathbf{w}$ denote any Nash equilibrium of the game defined by $(Q_1, \ldots, Q_R)$. We now investigate the efficiency loss of this system; that is, how much aggregate surplus is lost because the users attempt to "game" the system? To answer this question, we must compare the aggregate surplus $\sum_r U_r(d_r(\mathbf{w})) - C(\sum_r d_r(\mathbf{w}))$ obtained when the users fully evaluate the effect of their actions on the price, and the aggregate surplus $\sum_r U_r(d_r^S) - C(\sum_r d_r^S)$ obtained by choosing an allocation which maximizes aggregate surplus. The following theorem is the main result of this chapter: it states that the efficiency loss is no more than approximately 34%, and that this bound is essentially tight.

Theorem 3.8
Suppose that Assumptions 3.1-3.3 hold. Suppose also that $U_r(0) \ge 0$ for all r. If $\mathbf{d}^S$ is any optimal solution to SYSTEM, and $\mathbf{w}$ is any Nash equilibrium of the game defined by $(Q_1, \ldots, Q_R)$, then:

$$\sum_r U_r(d_r(\mathbf{w})) - C\left(\sum_r d_r(\mathbf{w})\right) \ge \left(4\sqrt{2} - 5\right)\left(\sum_r U_r(d_r^S) - C\left(\sum_r d_r^S\right)\right). \tag{3.25}$$

In other words, there is no more than approximately a 34% efficiency loss when users are price anticipating.

Furthermore, this bound is tight: for every $\delta > 0$, there exists a choice of R, a choice of (linear) utility functions U_r, $r = 1, \ldots, R$, and a (piecewise linear) price function p such that a Nash equilibrium $\mathbf{w}$ exists with:

$$\sum_r U_r(d_r(\mathbf{w})) - C\left(\sum_r d_r(\mathbf{w})\right) \le \left(4\sqrt{2} - 5 + \delta\right)\left(\sum_r U_r(d_r^S) - C\left(\sum_r d_r^S\right)\right). \tag{3.26}$$

Proof. The proof consists of a sequence of reductions:

1. We show that the worst case occurs when the utility function of each user is linear.

2. We show we may restrict attention to games where the total allocated Nash equilibrium rate is $f = 1$.

3. We compute the worst case choice of linear utility functions, for a fixed price function $p(\cdot)$ and total Nash equilibrium rate $f = 1$.

4. We show that it suffices to consider a special class of piecewise linear price functions.

5. Combining Steps 1-3, we compute the worst case efficiency loss by minimizing the ratio of Nash equilibrium aggregate surplus to maximal aggregate surplus, over the worst case choice of games with linear utility functions (from Step 2) and our restricted class of piecewise linear price functions (from Step 3).

Step 1: Show that we may assume without loss of generality that U_r is linear for each user r; i.e., $U_r(d_r) = \alpha_r d_r$, where $\alpha_1 = 1$ and $0 < \alpha_r \leq 1$ for $r > 1$. The proof of this claim is similar to the proof of Lemma 2.7. Let $\mathbf{d}^S$ denote any optimal solution to *SYSTEM*, and let $\mathbf{w}$ denote a Nash equilibrium, for an arbitrary collection of utility functions $(U_1, \ldots, U_R)$ satisfying the assumptions of the theorem. We let $\mathbf{d} = \mathbf{d}(\mathbf{w})$ denote the allocation vector at the Nash equilibrium. For each user r, we define a new utility function $\overline{U}_r(d_r) = \alpha_r d_r$, where $\alpha_r = U_r'(d_r)$; we know that $\alpha_r > 0$ by Assumption 3.1. Then observe that if we replace the utility functions $(U_1, \ldots, U_R)$ with the linear utility functions $(\overline{U}_1, \ldots, \overline{U}_R)$, the vector $\mathbf{w}$ remains a Nash equilibrium; this follows from the necessary and sufficient conditions of Proposition 3.6.

We first show that $\sum_r \alpha_r d_r - C(f) > 0$. To see this, note from (3.20) that $\alpha_r > p(f)$ for all r such that $d_r > 0$. Thus $\alpha_r d_r > d_r p(f)$ for such a user r, so $\sum_r \alpha_r d_r > f p(f) \geq C(f)$, by convexity (Assumption 3.2).

Next, we note that $\sum_r U_r(d_r^S) - C(\sum_r d_r^S) > 0$. This follows since U_r is strictly increasing and nonnegative, while $C'(0) = p(0) = 0$; thus if $\overline{d}_r$ is sufficiently small for all r, we will have $\sum_r U_r(\overline{d}_r) - C(\sum_r \overline{d}_r) > 0$, which implies $\sum_r U_r(d_r^S) - C(\sum_r d_r^S) > 0$ (since $\mathbf{d}^S$ is an optimal solution to *SYSTEM*).

Using concavity, we have for each r:

$$U_r(d_r^S) \leq U_r(d_r) + \alpha_r(d_r^S - d_r).$$

Thus we have:

$$\frac{\sum_r U_r(d_r) - C(\sum_r d_r)}{\sum_r U_r(d_r^S) - C(\sum_r d_r^S)} \geq \frac{\sum_r \big(U_r(d_r) - \alpha_r d_r\big) + \sum_r \alpha_r d_r - C(\sum_r d_r)}{\sum_r \big(U_r(d_r) - \alpha_r d_r\big) + \sum_r \alpha_r d_r^S - C(\sum_r d_r^S)}$$

$$\geq \frac{\sum_r \big(U_r(d_r) - \alpha_r d_r\big) + \sum_r \alpha_r d_r - C(\sum_r d_r)}{\sum_r \big(U_r(d_r) - \alpha_r d_r\big) + \max_{\overline{\mathbf{d}} \geq 0} \big(\sum_r \alpha_r \overline{d}_r - C(\sum_r \overline{d}_r)\big)}.$$

(Note all denominators above are positive, since we have shown that $\sum_r U_r(d_r^S) - C(\sum_r d_r^S) > 0$.) Since we assumed $U_r(0) \geq 0$, we have $U_r(d_r) - U_r'(d_r)d_r \geq 0$ by concavity; and since $0 < \sum_r \alpha_r d_r - C(f) \leq \max_{\overline{\mathbf{d}} \geq 0}(\sum_r \alpha_r \overline{d}_r - C(\sum_r \overline{d}_r))$, we have the inequality:

$$\frac{\sum_r U_r(d_r) - C(\sum_r d_r)}{\sum_r U_r(d_r^S) - C(\sum_r d_r^S)} \geq \frac{\sum_r \alpha_r d_r - C(\sum_r d_r)}{\max_{\overline{\mathbf{d}} \geq 0} \big(\sum_r \alpha_r \overline{d}_r - C(\sum_r \overline{d}_r)\big)}.$$

Now observe that the right hand side of the previous expression is the ratio of the Nash equilibrium aggregate surplus to the maximal aggregate surplus, when the utility functions are $(\overline{U}_1, \ldots, \overline{U}_R)$; since this ratio is no larger than the same ratio for the original utility functions $(U_1, \ldots, U_R)$, we can restrict attention to games where the utility function of each user is linear. Finally, by replacing α_r by $\alpha_r/(\max_s \alpha_s)$, and the cost function $C(\cdot)$ by $C(\cdot)/(\max_s \alpha_s)$, we may assume without loss of generality that $\max_r \alpha_r = 1$. Thus, by relabeling the users if necessary, we assume for the remainder of the proof that $U_r(d_r) = \alpha_r d_r$ for all r, where $\alpha_1 = 1$ and $0 < \alpha_r \leq 1$ for $r > 1$.

Before continuing, we observe that under these conditions, we have the following relation:

$$\max_{\overline{\mathbf{d}} \geq 0} \left(\sum_r \alpha_r \overline{d}_r - C\left(\sum_r \overline{d}_r\right)\right) = \max_{\overline{f} \geq 0} \left(\overline{f} - C(\overline{f})\right).$$

To see this, note that at any fixed value of $\overline{f} = \sum_r \overline{d}_r$, the left hand side is maximized by allocating the entire rate $\overline{f}$ to user 1. Thus, the ratio of Nash equilibrium aggregate surplus to maximal aggregate surplus becomes:

$$\frac{\sum_r \alpha_r d_r - C(\sum_r d_r)}{\max_{\overline{f} \geq 0} \big(\overline{f} - C(\overline{f})\big)}. \tag{3.27}$$

Note that the denominator is positive, since $C'(0) = p(0) = 0$; and further, the optimal solution in the denominator occurs at the unique value of $\overline{f} > 0$ such that $p(\overline{f}) = 1$.

Step 2: Show that we may restrict attention to games where the total allocated Nash equilibrium rate is $f = 1$. Fix a cost function C satisfying Assumption 3.2. Let $\mathbf{w}$ be a Nash equilibrium, and let $\mathbf{d} = \mathbf{d}(\mathbf{w})$ be the resulting allocation. Let $f = \sum_r d_r$ be the total allocated rate at the Nash equilibrium; note that $f > 0$ by Proposition 3.6. We now define a new price function $\hat{p}$ according to $\hat{p}(\hat{f}) = p(f\hat{f})$, and a new cost function $\hat{C}(\hat{f}) = \int_0^{\hat{f}} \hat{p}(z)\, dz$; note that $\hat{C}(\hat{f}) = C(f\hat{f})/f$. Then it is straightforward to check that $\hat{p}$ satisfies Assumption 3.2. We will use hats to denote the corresponding functions when the price function is $\hat{p}$: $\hat{\beta}^+(\hat{f})$, $\hat{\beta}^-(\hat{f})$, $\hat{d}_r(\mathbf{w})$, $\hat{f}(\mathbf{w})$, etc.

Define $\hat{w}_r = w_r/f$. Then we claim that $\hat{\mathbf{w}}$ is a Nash equilibrium when the price function is $\hat{p}$. First observe that:

$$\sum_r \hat{w}_r = \frac{\sum_r w_r}{f} = p(f) = \hat{p}(1).$$

Thus $\hat{f}(\hat{\mathbf{w}}) = 1$ by Proposition 3.1. Furthermore:

$$\hat{d}_r(\hat{\mathbf{w}}) = \frac{\hat{w}_r}{\hat{p}(\hat{f}(\hat{\mathbf{w}}))} = \frac{\hat{w}_r}{\hat{p}(1)} = \frac{w_r}{fp(f)} = \frac{d_r}{f}.$$

Finally, note that:

$$\frac{\partial^+ \hat{p}(1)}{\partial \hat{f}} = f\frac{\partial^+ p(f)}{\partial f},$$

from which we conclude that $\hat{\beta}^+(1) = \beta^+(f)$, and similarly $\hat{\beta}^-(1) = \beta^-(f)$. Recall that $\mathbf{w}$ is a Nash equilibrium when the price function is p; thus, if we combine the preceding conclusions and apply Proposition 3.6, we have that $\hat{\mathbf{w}}$ is a Nash equilibrium when the price function is $\hat{p}$, with total allocated rate 1 and allocation $\hat{\mathbf{d}} = \mathbf{d}/f$.

To complete the proof of this step, we note the following chain of inequalities:

$$\frac{\sum_r \alpha_r d_r - C(\sum_r d_r)}{\max_{\overline{f}\geq 0}\left(\overline{f} - C(\overline{f})\right)} = \frac{\sum_r \alpha_r \hat{d}_r - \hat{C}(1)}{\max_{\overline{f}\geq 0}\left(\overline{f}/f - C(\overline{f})/f\right)}$$

$$= \frac{\sum_r \alpha_r \hat{d}_r - \hat{C}(1)}{\max_{g\geq 0}\left(g - \hat{C}(g)\right)},$$

where we make the substitution $g = \overline{f}/f$. But now note that the right hand side is the ratio of Nash equilibrium aggregate surplus to maximal aggregate surplus for a game where the total Nash equilibrium allocated rate is equal to 1. Consequently, in computing the worst case efficiency loss, we may restrict our attention to games where the Nash equilibrium allocated rate is equal to 1.

Step 3: For a fixed price function p*, determine the instance of linear utility functions that minimizes Nash equilibrium aggregate surplus, for a fixed Nash equilibrium allocated rate* $f = \sum_r d_r = 1$. Note that fixing the price function p fixes the optimal aggregate surplus; thus minimizing the aggregate surplus at Nash equilibrium also yields the worst case efficiency loss.

We will optimize over the set of all games where users have linear utility functions (satisfying the conditions of Step 1), and where the total Nash equilibrium rate is $f = 1$. We use the necessary and sufficient conditions of Proposition 3.6. Note that fixing the price function p and the total rate $f > 0$ fixes the Nash equilibrium price, $p(1)$, as well as $\beta^+(1)$ and $\beta^-(1)$ (from the definition (3.18)); for notational convenience, we abbreviate $p = p(1)$, $C = C(1)$, $\beta^+ = \beta^+(1)$, and $\beta^- = \beta^-(1)$ for the duration of this step. Since $\alpha_1 = 1$, for a fixed value of R the game with linear utility functions that minimizes aggregate surplus is given by solving the following optimization problem (with unknowns $d_1, \dots, d_R, \alpha_2, \dots, \alpha_R$):

$$\text{minimize} \quad d_1 + \sum_{r=2}^{R} \alpha_r d_r - C \tag{3.28}$$

$$\text{subject to} \quad \alpha_r \left(1 - \beta^+ d_r\right) \le p, \quad r = 1, \dots, R; \tag{3.29}$$

$$\alpha_r \left(1 - \beta^- d_r\right) \ge p, \quad \text{if } d_r > 0, \; r = 1, \dots, R; \tag{3.30}$$

$$\sum_{r=1}^{R} d_r = 1; \tag{3.31}$$

$$0 < \alpha_r \le 1, \quad r = 2, \dots, R; \tag{3.32}$$

$$d_r \ge 0, \quad r = 1, \dots, R. \tag{3.33}$$

(Note that we have applied Proposition 3.6: if we solve the preceding problem and find an allocation $\mathbf{d}$ and coefficients $\boldsymbol{\alpha}$, then there exists a Nash equilibrium $\mathbf{w}$ with $\mathbf{d} = \mathbf{d}(\mathbf{w})$.) The objective function is the aggregate surplus given a Nash equilibrium allocation $\mathbf{d}$. The conditions (3.29)-(3.30) are equivalent to the Nash equilibrium conditions established in Proposition 3.6. The constraint (3.31) ensures that the total allocation made is equal to 1, and the constraint (3.32) follows from Step 1. The constraint (3.33) ensures the rate allocated to each user is nonnegative.

Our approach is to solve this problem through a sequence of reductions. We first show we may assume without loss of generality that the constraint (3.30) holds with equality for all users $r = 2, \dots, R$. The resulting problem is symmetric in the users $r = 2, \dots, R$; we next show that a feasible solution exists if and only if $1 - \beta^+ \le p < 1$ and R is sufficiently large, and we conclude using a convexity argument that $d_r = (f - d_1)/(R - 1)$ at an optimal solution. Finally, we show the worst case occurs in the limit where $R \to \infty$, and calculate the resulting Nash equilibrium aggregate surplus.

We first show that it suffices to optimize over all $(\boldsymbol{\alpha}, \mathbf{d})$ such that (3.30) holds with equality for $r = 2, \ldots, R$. Note that if $(\boldsymbol{\alpha}, \mathbf{d})$ is a feasible solution to (3.28)-(3.33), then from (3.30)-(3.33), and the fact that $0 < \beta^- < 1$, we conclude that $p < 1$. Now if $d_r > 0$ for some $r = 2, \ldots, R$, but the corresponding constraint in (3.30) does not hold with equality, we can reduce α_r until the constraint (3.30) becomes active; by this process we obtain a smaller value for the objective function (3.28). On the other hand, if $d_r = 0$ for some $r = 2, \ldots, R$, we can set $\alpha_r = p$; since $p < 1$, this preserves feasibility, but does not impact the term $\alpha_r d_r$ in the objective function (3.28). We can therefore restrict attention to feasible solutions for which:

$$\alpha_r = \frac{p}{1 - \beta^- d_r}, \quad r = 2, \ldots, R. \tag{3.34}$$

Having done so, observe that the constraint (3.32) that $\alpha_r \leq 1$ may be written as:

$$d_r \leq \frac{1-p}{\beta^-}, \quad r = 2, \ldots, R.$$

Finally, the constraint (3.32) that $\alpha_r > 0$ becomes redundant, as it is guaranteed by the fact that $d_r \leq 1$ (from (3.31)), $\beta^- < 1$ (by definition), and (3.34).

We now use the preceding observations to simplify the optimization problem (3.28)-(3.33) as follows:

$$\text{minimize} \quad d_1 + p \sum_{r=2}^{R} \frac{d_r}{1 - \beta^- d_r} - C \tag{3.35}$$

$$\text{subject to} \quad 1 - \beta^+ d_1 \leq p \leq 1 - \beta^- d_1; \tag{3.36}$$

$$\sum_{r=1}^{R} d_r = 1; \tag{3.37}$$

$$d_r \leq \frac{1-p}{\beta^-}, \quad r = 2, \ldots, R; \tag{3.38}$$

$$d_r \geq 0, \quad r = 1, \ldots, R. \tag{3.39}$$

The objective function (3.35) is equivalent to (3.28) upon substitution for α_r for $r = 2, \ldots, R$, from (3.34). We know $d_1 > 0$ when $p(f) < 1$ (from (3.29)-(3.30)); thus the constraint (3.36) is equivalent to the constraints (3.29)-(3.30) for user 1 with $d_1 > 0$. The constraint (3.29) for $r > 1$ is redundant and eliminated, since (3.30) holds with equality for $r > 1$. The constraint (3.37) is equivalent to the allocation constraint (3.31); and the constraint (3.38) ensures $\alpha_r \leq 1$, as required in (3.32).

We first note that for a feasible solution to (3.35)-(3.39) to exist, we must have $1 - \beta^+ \leq p < 1$. We have already shown that we must have $p < 1$ if a feasible solution exists. Furthermore, from (3.36) we observe that the smallest feasible value of d_1 is $d_1 =$

$(1-p)/\beta^+$. We require $d_1 \le 1$ from (3.37) and (3.39), so we must have $(1-p)/\beta^+ \le 1$, which yields the restriction that $1 - \beta^+ \le p$. Thus, there only exist Nash equilibria with total rate 1 and price p if:

$$1 - \beta^+ \le p < 1. \tag{3.40}$$

We will assume for the remainder of this step that (3.40) is satisfied.

We note that if $\overline{\mathbf{d}} = (\overline{d}_1, \dots, \overline{d}_R)$ is a feasible solution to (3.35)-(3.39) with R users, then letting $\overline{d}_{R+1} = 0$, the vector $(\overline{d}_1, \dots, \overline{d}_{R+1})$ is a feasible solution to (3.35)-(3.39) with $R+1$ users, and with the same objective function value (3.35) as $\overline{\mathbf{d}}$. Thus the minimal objective function value cannot increase as R increases, so the worst case efficiency loss occurs in the limit where $R \to \infty$.

We now solve (3.35)-(3.39) for a fixed feasible value of d_1. From the constraints (3.37)-(3.38), we observe that a feasible solution to (3.35)-(3.39) exists if and only if the following condition holds in addition to (3.40):

$$d_1 + (R-1)\left(\frac{1-p}{\beta^-}\right) \ge 1. \tag{3.41}$$

In this case, the following symmetric solution is feasible:

$$d_r = \frac{1-d_1}{R-1}, \quad r = 2, \dots, R. \tag{3.42}$$

Furthermore, since the objective function is strictly convex and symmetric in the variables $d_2, \dots, d_R$, and the feasible region is convex, the symmetric solution (3.42) must be optimal.

If we substitute the optimal solution (3.42) into the objective function (3.35) and take the limit as $R \to \infty$, then the constraint (3.41) is vacuously satisfied, and the objective function becomes $d_1 + p(1-d_1) - C$. Since we have shown that $p < 1$, the worst case occurs at the smallest feasible value of d_1; from (3.36), this value is:

$$d_1 = \frac{1-p}{\beta^+}. \tag{3.43}$$

The resulting worst case Nash equilibrium aggregate surplus is:

$$p + \frac{(1-p)^2}{\beta^+} - C.$$

To complete the proof of the theorem, we will consider the ratio of this Nash equilibrium aggregate surplus to the maximal aggregate surplus; we denote this ratio by

$F(p)$, as a function of the price function $p(\cdot)$:

$$F(p) = \frac{p(1) + (1 - p(1))^2/\beta^+(1) - C(1)}{\max_{f \geq 0} (f - C(f))}. \tag{3.44}$$

Note that henceforth, the scalar p used throughout Step 3 will be denoted $p(1)$, and we return to denoting the price function by p. Thus $F(p)$ as defined in (3.44) is a function of the entire price function $p(\cdot)$.

For completeness, we summarize in the following lemma an intermediate tightness result which will be necessary to prove the tightness of the bound in the theorem.

Lemma 3.9 *Suppose that Assumptions 3.2-3.3 hold. Then there exists $R > 0$ and a choice of linear utility functions $U_r(d_r) = \alpha_r d_r$, where $\alpha_1 = \max_s \alpha_s = 1$, with total Nash equilibrium rate 1, if and only if* (3.40) *is satisfied, i.e.:*

$$1 - \beta^+(1) \leq p(1) < 1. \tag{3.45}$$

In this case, given $\delta > 0$, there exists $R > 0$ and a collection of R users where user r has utility function $U_r(d_r) = \alpha_r d_r$, such that $\mathbf{d}$ is a Nash equilibrium allocation with $\sum_r d_r = 1$, and:

$$\frac{\sum_r \alpha_r d_r - C(1)}{\max_{\mathbf{d} \geq 0} \left(\sum_r \alpha_r d_r - C(\sum_r d_r) \right)} \leq F(p) + \delta. \tag{3.46}$$

Proof of Lemma. The proof follows from Step 3. We have shown that if there exists a Nash equilibrium with total rate 1, then (3.45) must be satisfied. Conversely, if (3.45) is satisfied, we proceed as follows: define d_1 according to (3.43); choose R large enough that (3.41) is satisfied; define d_r according to (3.42); and then define α_r according to (3.34) with $p = p(1)$. Then it follows that $(\mathbf{d}, \boldsymbol{\alpha})$ is a feasible solution to (3.28)-(3.33), which (by Proposition 3.6) guarantees there exists a Nash equilibrium with total allocated rate equal to 1.

The bound in (3.46) then follows by the proof of Step 3. □

The remainder of the proof amounts to minimizing the worst case ratio of Nash equilibrium aggregate surplus to maximal aggregate surplus, over all valid choices of p—that is, price functions p such that at least one choice of linear utility functions satisfying the conditions of Step 1 leads to a Nash equilibrium with total rate 1. By Lemma 3.9, all such functions p are characterized by the constraint (3.45). We will minimize $F(p)$, given by (3.44), over all choices of p satisfying (3.45).

Step 4: Show that in minimizing $F(p)$ over p satisfying (3.45)*, we may restrict attention to functions p satisfying the following conditions:*

$$p(f) = \begin{cases} af, & 0 \le f \le 1; \\ a + b(f-1), & f \ge 1; \end{cases} \tag{3.47}$$

$$0 < a \le b; \tag{3.48}$$

$$\frac{1}{a+b} \le 1 < \frac{1}{a}. \tag{3.49}$$

Observe that p as defined in (3.47)-(3.49) is a convex, strictly increasing, piecewise linear function with two parts: an initial segment which increases at slope $a > 0$, and a second segment which increases at slope $b \ge a$. In particular, such a function satisfies Assumption 3.2. Furthermore, we have $\partial^+ p(1)/\partial f = b$, so that $\varepsilon^+(1) = b/a$. This implies $\beta^+(1) = b/(a+b)$; thus, multiplying through (3.49) by a yields (3.45).

To verify the claim of Step 4, we consider any p such that (3.45) holds. We define a new price function $\overline{p}$ as follows:

$$\overline{p}(f) = \begin{cases} fp(1), & 0 \le f \le 1; \\ \\ p(1) + \dfrac{\partial^+ p(1)}{\partial f}(f-1), & f \ge 1. \end{cases} \tag{3.50}$$

(See Figure 3-2 for an illustration.) Let $a = p(1)$, and let $b = \partial^+ p(1)/\partial f$. Then $a > 0$; and since $p(0) = 0$, we have $\partial^+ p(1)/\partial f \ge p(1)$ by convexity of p, so that $b \ge a$. Furthermore, since $p(1) < 1$ from (3.45), we have $1/a > 1$. Finally, we have:

$$\frac{1}{a+b} = \frac{1}{p(1)}(1 - \beta^+(1)) \le 1,$$

where the equality follows from the definition of $\beta^+(1)$ and the inequality follows from (3.45). Thus $\overline{p}$ satisfies (3.47)-(3.49). Observe also that $\overline{p}(1) = p(1)$, and $\partial^+ \overline{p}(1)/\partial f = \partial^+ p(1)/\partial f$, so that $\overline{\beta}^+(1) = \beta^+(1)$.

We now show that $F(\overline{p}) \le F(p)$. As an intermediate step, we define a new price function $\hat{p}(\cdot)$ as follows:

$$\hat{p}(f) = \begin{cases} p(f), & 0 \le f \le 1; \\ \overline{p}(f), & f \ge 1. \end{cases}$$

Of course, $\hat{p}(1) = p(1)$ and $\partial^+ \hat{p}(1)/\partial f = \partial^+ \overline{p}(1)/\partial f = \partial^+ p(1)/\partial f$, so that (3.45) is satisfied for $\hat{p}$. Let $\hat{C}(f) = \int_0^f \hat{p}(z)\ dz$ denote the cost function associated with $\hat{p}(\cdot)$. Observe that (by convexity of p), we have for all f that $\hat{p}(f) \le p(f)$, so that $\hat{C}(f) \le$

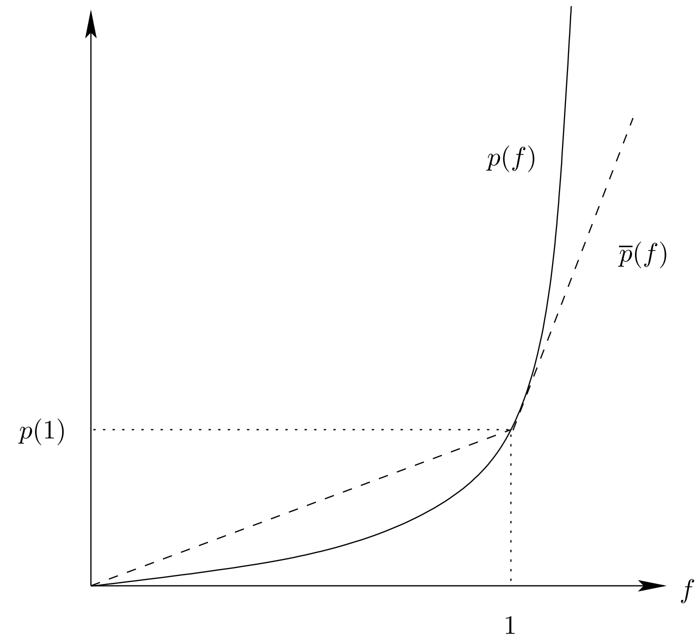

Figure 3-2. Proof of Theorem 3.8, Step 4: Given a price function p (solid line) and Nash equilibrium rate 1, a new price function $\overline{p}$ (dashed line) is defined according to (3.50).

$C(f)$. Thus:

$$\max_{f\geq 0}(f - \hat{C}(f)) \geq \max_{f\geq 0}(f - C(f)) > 0.$$

Furthermore, $\hat{C}(1) = C(1)$ trivially; hence we conclude $F(\hat{p}) \leq F(p)$.

Next, we let $\overline{C}(f) = \int_0^f \overline{p}(z)\ dz$ denote the cost function associated with $\overline{p}(\cdot)$. By convexity of p, we know $\overline{p}(1) \geq p(1)$ for $0 \leq f \leq 1$; thus $\overline{C}(f) \geq C(f)$ in that region. We let $\Delta = \overline{C}(1) - C(1) \geq 0$. Then we have the following relationship:

$$F(\hat{p}) = \frac{p(1) + (1 - p(1))^2/\beta^+(1) - C(1)}{\max_{f\geq 0}\left(f - \hat{C}(f)\right)} \tag{3.51}$$

$$\geq \frac{p(1) + (1 - p(1))^2/\beta^+(1) - (C(1) + \Delta)}{\max_{f\geq 0}\left(f - (\hat{C}(f) + \Delta)\right)} \tag{3.52}$$

$$= F(\overline{p}). \tag{3.53}$$

The last equality follows by observing that since $\hat{p}(1) = p(1) < 1$, the optimal solution to $\max_{f\geq 0}(f - \hat{C}(f))$ occurs at $\hat{f}^S > 1$ where $\hat{p}(\hat{f}^S) = 1$; and at all points $f \geq 1$, we have the relationship $\hat{C}(f) + \Delta = \overline{C}(f)$. Combining the preceding results, we have $F(p) \geq F(\overline{p})$, as required.

Step 5: The minimum value of $F(p)$ over all p satisfying (3.47)-(3.49) *is* $4\sqrt{2}-5$. Our first step is to show that given p satisfying (3.47)-(3.49), $F(p)$ is given by:

$$F(p) = \frac{\frac{1}{2}a + \left(1+\frac{a}{b}\right)(1-a)^2}{1-\frac{1}{2}a+\frac{1}{2}\frac{(1-a)^2}{b}} = \frac{ab+2(a+b)(1-a)^2}{2b-ab+(1-a)^2}. \tag{3.54}$$

The numerator results by simplifying the numerator of (3.44), when p takes the form described by (3.47)-(3.49). To arrive at the denominator, we note that the optimal solution to $\max_{f\geq 0}(f - C(f))$ occurs at f^S satisfying $p(f^S) = 1$. Since $a < 1$, we must have $f^S > 1$ and:

$$a + b(f^S - 1) = 1.$$

Simplifying, we find:

$$f^S = 1 + \frac{1-a}{b}. \tag{3.55}$$

The expression $f^S - C(f^S)$, upon simplification, becomes the denominator of (3.54), as required.

Fix a and b such that $0 < a \leq b$, and $1/(a+b) \leq 1 < 1/a$, and define p as in (3.47). We note here that the constraints $0 < a \leq b$ and $1/(a+b) \leq 1 < 1/a$ may be equivalently rewritten as $0 < a < 1$, and $\max\{a, 1-a\} \leq b$. Let $H(a,b) = F(p)$; from (3.54), note that for fixed a, $H(a,b)$ is a ratio of two affine functions of b, and thus the minimal value of $H(a,b)$ is achieved either when $b = \max\{a, 1-a\}$ or as $b \to \infty$. Define $H_1(a) = H(a,b)|_{\max\{a,1-a\}}$, and $H_2(a) = \lim_{b\to\infty} H(a,b)$. Then:

$$H_1(a) = \begin{cases} H(a,b)|_{b=1-a} & = \dfrac{2-a}{3-2a}, \qquad \text{if } 0 < a \leq 1/2; \\ \\ H(a,b)|_{b=a} & = a^2 + 4a(1-a)^2, \qquad \text{if } 1/2 \leq a < 1; \end{cases} \tag{3.56}$$

$$H_2(a) = \lim_{b\to\infty} H(a,b) = \frac{a + 2(1-a)^2}{2-a}. \tag{3.57}$$

We now minimize $H_1(a)$ and $H_2(a)$ over $0 < a < 1$. Over $0 < a \leq 1/2$, the minimum value of $H_1(a)$ is $2/3$, achieved as $a \to 0$. Over $1/2 \leq a < 1$, the minimum value of $H_1(a)$ is $20/27$, achieved at $a = 2/3$. Finally, over $0 < a < 1$, the minimum value of $H_2(a)$ is $4\sqrt{2}-5$, achieved at $a = 2-\sqrt{2}$. Since $\min\{2/3, 20/27, 4\sqrt{2}-5\} = 4\sqrt{2}-5$, we conclude that the minimal value of $F(p)$ over all p satisfying (3.47)-(3.49) is equal to $4\sqrt{2}-5$. This completes the proof of (3.25).

We now show that this lower bound is tight. Fix $\delta > 0$. The preceding argument shows that the worst case occurs for price functions satisfying (3.47)-(3.49), where $a =$

$2-\sqrt{2}$ and $b \to \infty$. For fixed $b \geq a = 2-\sqrt{2}$, let p_b be the associated price function defined according to (3.47). Then we have established that:

$$\lim_{b\to\infty} F(p_b) = 4\sqrt{2} - 5.$$

From Lemma 3.9, we know there exists γ_b such that $\gamma_b < F(p_b) + \delta/2$, and where γ_b is the ratio of Nash equilibrium aggregate surplus to maximal aggregate surplus for some game with price function p_b and total allocated rate 1 at the Nash equilibrium. We thus have:

$$\lim_{b\to\infty} \gamma_b = \lim_{b\to\infty} F(p_b) + \delta/2 = 4\sqrt{2} - 5 + \delta/2.$$

Thus for b sufficiently large, we will have $\gamma_b < 4\sqrt{2} - 5 + \delta$, establishing (3.26). □

The preceding theorem shows that in the worst case, aggregate surplus falls by no more than approximately 34% when users are able to anticipate the effects of their actions on the price of the link. Furthermore, this bound is essentially tight. In fact, from the proof of the theorem we see that this ratio is achieved via a sequence of games where:

1. The price function p has the form given by (3.47)-(3.49), with $a = 2-\sqrt{2}$, $b \to \infty$, and $f = 1$;

2. The number of users becomes large ($R \to \infty$); and

3. User 1 has linear utility with $U_1(d_1) = d_1$, and all other users r have linear utility with $U_r(d_r) = \alpha_r d_r$, where $\alpha_r \approx p(1) = 2 - \sqrt{2}$.

The last item follows by substituting the solution (3.42) in (3.34), and taking the limit as $R \to \infty$. (Note that formally, we must take care that the limits of $R \to \infty$ and $b \to \infty$ are taken in the correct order; in particular, in the proof we first have $R \to \infty$, and then $b \to \infty$.)

It is interesting to note that the worst case is obtained by considering instances where the price function is becoming steeper and steeper at the Nash equilibrium rate 1, since $b \to \infty$. This forces the system optimal rate f^S to approach the Nash equilibrium rate $f = 1$, as we observe from (3.55); nevertheless, the shortfall between the Nash equilibrium aggregate surplus and the maximal aggregate surplus approaches 34%.

■ 3.3 Inelastic Supply vs. Elastic Supply

In this section we briefly compare the model of this chapter (allocation of a resource in elastic supply) with the model of Chapter 2 (allocation of a resource in inelastic supply). In Section 2.1, a model is considered with a single link having exactly C units of rate available to allocate among the users. As in the model of this chapter, user r submits a bid w_r. The link manager then sets a price $\mu = \sum_r w_r/C$; and user r receives an allocation d_r given by:

$$d_r = \begin{cases} 0, & \text{if } w_r = 0; \\ \dfrac{w_r}{\mu}, & \text{if } w_r > 0. \end{cases}$$

Formally, the space of parametrized *demand functions* available to the users are the same in both models: $D(\mu, w_r) = w_r/\mu$. (See Sections 2.1 and 3.1 for more on this point.) As in this chapter, in Chapter 2 the payoff to user r is $U_r(d_r) - w_r$. It is shown in Section 2.2 that when users are price anticipating and the link supply is inelastic, the efficiency loss is at most 25%.

Intuitively, we would like to model a system with an inelastic supply C by a cost function which is zero for $0 \leq f < C$, and infinite for $f > C$. Formally, we show in this section that if the price function is given by $p(f) = af^B$ for $a \geq 0$ and $B \geq 1$, then as $B \to \infty$ the worst case ratio of Nash equilibrium aggregate surplus to maximal aggregate surplus approaches 75%—the same value obtained in Theorem 2.6. While this does not formally establish the result of Theorem 2.6, the limit is intuitively plausible, because as the exponent B increases, the price function p and associated cost function begin to resemble an inelastic capacity constraint with $C = 1$: for $f < 1$, $f^B \to 0$ as $B \to \infty$; and for $f > 1$, $f^B \to \infty$ as $B \to \infty$.

Theorem 3.10
Suppose that Assumptions 3.1-3.3 hold. Suppose also that $U_r(0) \geq 0$ for all r, and that $p(f) = af^B$ with $a > 0$ and $B \geq 1$. Define the function $g(B)$ by:

$$g(B) = \left(\frac{B+1}{2B+1}\right)^{1/B} \left(\frac{(B+1)(3B+2)}{(2B+1)^2}\right). \tag{3.58}$$

If $\mathbf{d}^S$ is any optimal solution to SYSTEM, and $\mathbf{w}$ is any Nash equilibrium of the game defined by $(Q_1, \dots, Q_R)$, then:

$$\sum_r U_r(d_r(\mathbf{w})) - C\left(\sum_r d_r(\mathbf{w})\right) \geq g(B)\left(\sum_r U_r(d_r^S) - C\left(\sum_r d_r^S\right)\right). \tag{3.59}$$

The function $g(B)$ is strictly increasing, with $g(B) \to 3/4$ as $B \to \infty$; see Figure 3-3.

Furthermore, the bound (3.59) *is tight: for fixed $B \geq 1$, for every $\delta > 0$, there exists a choice of R and a choice of (linear) utility functions U_r, $r = 1, \dots, R$, such that a Nash*

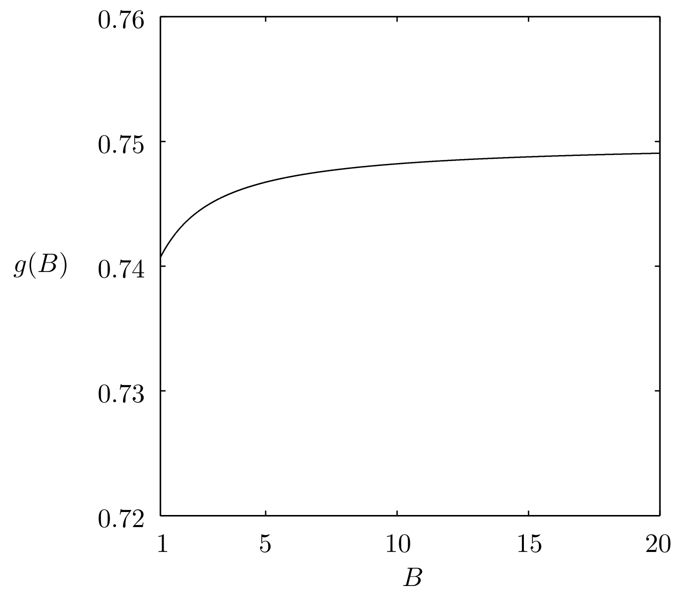

Figure 3-3. The function $g(B)$ in Theorem 3.10: The function $g(B)$ is defined for $B \geq 1$ in (3.58), and gives the worst case efficiency loss when the price function is $p(f) = af^B$ with $a > 0$. Note that $g(B)$ is strictly increasing, with $g(1) = 20/27$ and $g(B) \to 3/4$ as $B \to \infty$.

equilibrium $\mathbf{w}$ *exists with:*

$$\sum_r U_r(d_r(\mathbf{w})) - C\left(\sum_r d_r(\mathbf{w})\right) \leq (g(B) + \delta)\left(\sum_r U_r(d_r^S) - C\left(\sum_r d_r^S\right)\right). \quad (3.60)$$

Proof. We follow the proof of Theorem 3.8. Steps 1-4 follow as in that proof, provided we can show that two scalings of the function $p(\cdot)$ do not affect our result—in Step 1, where we replace $p(\cdot)$ by $p(\cdot)/\max_r \alpha_r$, and in Step 2, where we replace $p(\cdot)$ by $p(f\cdot)$, where f is the Nash equilibrium rate. Indeed, both these scalings remain valid, since the rescaled price function is still a monomial with the same exponent as p, but a different constant coefficient. In particular, we may continue to restrict attention to the special case where $U_r(d_r) = \alpha_r d_r$, with $\max_r \alpha_r = \alpha_1 = 1$, and where the total Nash equilibrium allocated rate is 1.

From Steps 1-4 of the proof of Theorem 3.8, we must minimize $F(p)$, defined in (3.44), for all choices of p such that (3.45) is satisfied, i.e., such that:

$$1 - \beta^+(1) \leq p(1) < 1.$$

For $p(f) = af^B$, we have $\beta^+(f) = B/(1+B)$; and thus we require:

$$\frac{1}{1+B} \leq a < 1. \tag{3.61}$$

Note that at the maximal aggregate surplus, $p(f^S) = a(f^S)^B = 1$ implies that $f^S = a^{-1/B}$. Furthermore, $C(f) = af^{B+1}/(B+1)$ for $f \geq 0$. Thus $f^S - C(f^S)$ is given by:

$$f^S - C(f^S) = \left(\frac{B}{B+1}\right)\left(\frac{1}{a}\right)^{1/B}.$$

From (3.44), we conclude that $F(p)$ is given by:

$$F(p) = \frac{a + (1-a)^2(1+1/B) - a/(B+1)}{\left(\frac{B}{B+1}\right)\left(\frac{1}{a}\right)^{1/B}}.$$

We now minimize $F(p)$ over the set of a satisfying (3.61). We begin by differentiating $F(p)$ with respect to a, and setting the derivative to zero; simplifying, this yields the following equation:

$$Ba + \left(1 + \frac{1}{B}\right)\left((2B+1)a^2 - 2(B+1)a + 1\right) = 0.$$

This equation is quadratic in a, and has two solutions a_1 and a_2: $a_1 = 1/(B+1)$, and $a_2 = (B+1)/(2B+1)$. Both solutions satisfy (3.61). Let $p_1(f) = a_1 f^B$, and $p_2(f) = a_2 f^B$. We have:

$$F(p_1) = \left(\frac{1}{B+1}\right)^{1/B}\left(\frac{B+2}{B+1}\right); \quad F(p_2) = \left(\frac{B+1}{2B+1}\right)^{1/B}\left(\frac{(B+1)(3B+2)}{(2B+1)^2}\right).$$

To minimize $F(p)$ over a satisfying (3.61), we need also to check the endpoint where $a = 1$. If $p = f^B$, we find $F(p) = 1$; since $F(p_1), F(p_2) \leq 1$ from the definition of $F(p)$, the minimum value is achieved at either p_1 or p_2.

For $B \geq 1$, we define $g_1(B) = F(p_1)$, and $g_2(B) = F(p_2)$. We need the following technical lemma.

Lemma 3.11 *The functions $g_1(B)$ and $g_2(B)$ are strictly increasing for $B \geq 1$. Furthermore, $g_1(B) \geq 3/4$ for $B \geq 1$, while $\lim_{B\to\infty} g_2(B) = 3/4$.*

Proof. We begin by noting that $g_1(1) = 3/4$. Let $\hat{g}_1(B) = \ln(g_1(B))$; it suffices to show that $\hat{g}_1(B)$ is strictly increasing for $B \geq 1$. Differentiating $\hat{g}_1$ yields:

$$\hat{g}_1'(B) = \frac{(B+2)\ln(B+1) - 2B}{(B+2)B^2}.$$

It suffices to check that $h_1(B) > 0$, where:

$$h_1(B) = (B+2)\ln(B+1) - 2B.$$

We have $h_1(1) = 3\ln 2 - 2 > 0$; $h_1'(1) = \ln 2 - 1/2 > 0$; and $h_1''(B) = B/(B+1)^2 > 0$. This implies $h_1(B) > 0$ for all $B \geq 1$, so g_1 is strictly increasing for $B \geq 1$.

Next we consider $g_2(B)$. Note first that $g_2(1) = 20/27$. Furthermore, as $B \to \infty$, $((B+1)/(2B+1))^{1/B} \to 1$, and $(B+1)(3B+2)/(2B+1)^2 \to 3/4$. Thus $g_2(B) \to 3/4$ as $B \to \infty$.

Finally, let $\hat{g}_2(B) = \ln(g_2(B))$; it suffices to show $\hat{g}_2$ is strictly increasing for $B \geq 1$. Differentiating $\hat{g}_2(B)$ yields:

$$\hat{g}_2'(B) = \frac{(3B+2)\ln\left(\frac{2B+1}{B+1}\right) - 2B}{(3B+2)B^2}.$$

As above, it suffices to check that $h_2(B) > 0$, where:

$$h_2(B) = (3B+2)\ln\left(\frac{2B+1}{B+1}\right) - 2B.$$

We have $h_2(1) = 5\ln(3/2) - 2 > 0$; $h_2'(1) = 3\ln(3/2) - 7/6 > 0$; and $h_2''(1) = B/[(B+1)^2(2B+1)^2] > 0$. Thus $h_2(B) > 0$ for all $B \geq 1$, which implies g_2 is strictly increasing for $B \geq 1$. □

From the previous lemma, we conclude that the minimum value of $F(p)$ over $p = af^B$ satisfying (3.61) is given by $g_2(B)$; this establishes (3.59). As in Theorem 3.8, by construction this bound is tight, so (3.60) holds as well. □

The preceding theorem shows that for a particular sequence of price functions which approach a "hard" capacity constraint, the efficiency loss gradually decreases from $7/27$ (at $B = 1$) to $1/4$ (as $B \to \infty$). In the limit as $B \to \infty$, we recover the same efficiency loss as in Theorem 2.6. However, while we have demonstrated such a limit holds as long as the price functions are monomials, there remains an open question: if the price functions "converge" (in an appropriate sense) to a hard capacity constraint, under what conditions does the efficiency loss also converge to $1/4$? It is straightforward to check that such a limit cannot always hold. For example, consider price functions p of the form specified in (3.47)-(3.49). Using the expression for $F(p)$ given in (3.54), it is possible to show that by first taking $b \to \infty$, and then taking $a \to 0$, the worst case efficiency loss approaches zero; see (3.57).

3.4 General Networks

In this section we will consider an extension of the single link model to general networks. We adopt exactly the same network model as Section 2.4. We consider a network consisting of J links, or *resources*, numbered $1, \ldots, J$. As before, a set of users numbered $1, \ldots, R$, shares this network of resources. We assume there exists a set of paths through the network, numbered $1, \ldots, P$. By an abuse of notation, we will use J, R, and P to also denote the sets of resources, users, and paths, respectively. Each path $q \in P$ uses a subset of the set of resources J; if resource j is used by path q, we will denote this by writing $j \in q$. (Note that we now denote a path by q rather than p, as in Chapter 2; this change is made to avoid confusion with the price function p.) Each user $r \in R$ has a collection of paths available through the network; if path q serves user r, we will denote this by writing $q \in r$. We will assume without loss of generality that paths are uniquely identified with users, so that for each path q there exists a unique user r such that $q \in r$. (There is no loss of generality because if two users share the same path, that is captured in our model by creating two paths which use exactly the same subset of resources.) For notational convenience, we note that the resources required by individual paths are captured by the *path-resource incidence matrix* $\mathbf{A}$, defined by:

$$A_{jq} = \begin{cases} 1, & \text{if } j \in q \\ 0, & \text{if } j \notin q. \end{cases}$$

Furthermore, we can capture the relationship between paths and users by the *path-user incidence matrix* $\mathbf{H}$, defined by:

$$H_{rq} = \begin{cases} 1, & \text{if } q \in r \\ 0, & \text{if } q \notin r. \end{cases}$$

Note that by our assumption on paths, for each path q we have $H_{rq} = 1$ for exactly one user r.

Let $y_q \geq 0$ denote the rate allocated to path q, and let $d_r = \sum_{q \in r} y_q \geq 0$ denote the rate allocated to user r; using the matrix $\mathbf{H}$, we may write the relation between $\mathbf{d} = (d_r, r \in R)$ and $\mathbf{y} = (y_q, q \in P)$ as $\mathbf{Hy} = \mathbf{d}$. Furthermore, if we let f_j denote the total rate on link j, we must have:

$$\sum_{q: j \in q} y_q = f_j, \quad j \in J.$$

Using the matrix $\mathbf{A}$, we may write this constraint as $\mathbf{Ay} = \mathbf{f}$.

We continue to assume that user r receives a utility $U_r(d_r)$ from an allocated rate d_r, and that each link j incurs a cost $C_j(f_j)$ when the total allocated rate at link j is f_j. We make the following assumptions regarding the utility functions and cost functions.

Assumption 3.5
For each r, the utility function $U_r(d_r)$ is concave, nondecreasing, and continuous over the domain $d_r \geq 0$.

Assumption 3.6
For each j, there exists a continuous, convex, strictly increasing function $p_j(f_j)$ over $f_j \geq 0$ with $p_j(0) = 0$, such that for $f_j \geq 0$:

$$C_j(f_j) = \int_0^{f_j} p_j(z)dz.$$

Thus $C_j(f_j)$ is strictly convex and increasing.

Assumption 3.5 is similar to Assumption 3.1, but we no longer require that U_r be strictly increasing or differentiable. (Thus Assumption 3.5 is identical to Assumption 2.2.) Assumption 3.6 is identical to Assumption 3.2, for each link j. We emphasize here that while we consider a more general class of utility functions under Assumption 3.5, we do not resort to an extended game in this section as we did in Section 2.4.1. For this reason, in the single link setting, the results of this section are generalizations of the results of Sections 3.1 and 3.2.

The natural generalization of the problem *SYSTEM* to a network context is given by the following optimization problem:

SYSTEM:

$$\text{maximize} \quad \sum_r U_r(d_r) - \sum_j C_j(f_j) \tag{3.62}$$

$$\text{subject to} \quad \mathbf{Ay} = \mathbf{f}; \tag{3.63}$$

$$\mathbf{Hy} = \mathbf{d}; \tag{3.64}$$

$$y_q \geq 0, \quad q \in P. \tag{3.65}$$

We continue to refer to the objective function (3.62) as the *aggregate surplus* (see Section 1.1). Since the objective function is continuous and U_r grows at most linearly while C_j grows superlinearly, an optimal solution $\mathbf{y}$ exists. Since the feasible region is convex and the cost functions C_j are each strictly convex, the optimal vector $\mathbf{f} = \mathbf{Ay}$ is uniquely defined (though $\mathbf{y}$ need not be unique). In addition, if the functions U_r are strictly concave, then the optimal vector $\mathbf{d} = \mathbf{Hy}$ is uniquely defined as well. As in the previous development, we will use the optimal solution to *SYSTEM* as a benchmark for the outcome of the network game.

We now define the resource allocation mechanism for this network setting, following the development of Section 2.4. The natural extension of the single link model is defined as follows. Each user r submits a *bid* w_{jr} for each resource j; this defines a strategy for user r given by $\mathbf{w}_r = (w_{jr}, j \in J)$, and a composite strategy vector given by $\mathbf{w} = (\mathbf{w}_1, \ldots, \mathbf{w}_R)$. We then assume that each link takes these bids as input, and uses the pricing scheme developed in Section 3.1. This is formalized in the following assumption, which is a direct analogue of Assumption 3.3 for each link j.

Assumption 3.7
For all $\mathbf{w} \geq 0$, at each link j the aggregate rate $f_j(\mathbf{w})$ is the unique solution f_j to:

$$\sum_r w_{jr} = f_j p_j(f_j). \tag{3.66}$$

Furthermore, for each r, $x_{jr}(\mathbf{w})$ is given by:

$$x_{jr}(\mathbf{w}) = \begin{cases} 0, & \text{if } w_{jr} = 0; \\ \dfrac{w_{jr}}{p_j(f_j(\mathbf{w}))}, & \text{if } w_{jr} > 0. \end{cases} \tag{3.67}$$

We define the vector $\mathbf{x}_r(\mathbf{w})$ by:

$$\mathbf{x}_r(\mathbf{w}) = (x_{jr}(\mathbf{w}), j \in J).$$

Now given any vector $\overline{\mathbf{x}}_r = (\overline{x}_{jr}, j \in J)$, we define $d_r(\overline{\mathbf{x}}_r)$ to be the optimal objective value of the following optimization problem:

$$\text{maximize} \quad \sum_{q \in r} y_q \tag{3.68}$$

$$\text{subject to} \quad \sum_{q \in r : j \in q} y_q \leq \overline{x}_{jr}, \quad j \in J; \tag{3.69}$$

$$y_q \geq 0, \quad q \in r. \tag{3.70}$$

(Note this is identical to the definition in (2.43)-(2.45).) Given the strategy vector $\mathbf{w}$, we then define the rate allocated to user r as $d_r(\mathbf{x}_r(\mathbf{w}))$. Thus user r is allocated the maximum possible rate possible, given that each link j has granted him rate $x_{jr}(\mathbf{w})$.

Define the notation $\mathbf{w}_{-r} = (\mathbf{w}_1, \ldots, \mathbf{w}_{r-1}, \mathbf{w}_{r+1}, \ldots, \mathbf{w}_R)$. Based on the definition of $d_r(\mathbf{x}_r(\mathbf{w}))$ above, the payoff to user r is given by:

$$T_r(\mathbf{w}_r; \mathbf{w}_{-r}) = U_r\big(d_r(\mathbf{x}_r(\mathbf{w}))\big) - \sum_j w_{jr}. \tag{3.71}$$

A *Nash equilibrium* of the game defined by $(T_1, \ldots, T_R)$ is a vector $\mathbf{w} \geq 0$ such that for all r:

$$T_r(\mathbf{w}_r; \mathbf{w}_{-r}) \geq T_r(\overline{\mathbf{w}}_r; \mathbf{w}_{-r}), \quad \text{for all } \overline{\mathbf{w}}_r \geq 0. \tag{3.72}$$

As in the development of Section 3.1.2, the following proposition plays a key role in demonstrating existence of a Nash equilibrium. The proof is identical to the proof of Proposition 3.3, and is omitted.

Proposition 3.12
Suppose that Assumptions 3.5-3.7 hold. Then for each j and each r: (1) $x_{jr}(\mathbf{w})$ is a continuous function of $\mathbf{w}$; and (2) for any $\mathbf{w}_{-r} \geq 0$, $x_{jr}(\mathbf{w})$ is strictly increasing and concave in $w_{jr} \geq 0$, and $x_{jr}(\mathbf{w}) \to \infty$ as $w_{jr} \to \infty$.

As in Proposition 3.4, the following proposition gives existence of a Nash equilibrium for the game defined by $(T_1, \dots, T_R)$. We note here that when supply is elastic, we do not have any discontinuity in the payoff function T_r of user r; and thus we do not require an extended game to guarantee existence of a Nash equilibrium, as was developed in Section 2.4.1.

Proposition 3.13
Suppose that Assumptions 3.5-3.7 hold. Then there exists a Nash equilibrium $\mathbf{w}$ for the game defined by $(T_1, \dots, T_R)$.

Proof. The proof follows the proof of Proposition 3.4. The only step which requires modification is to show that the payoff T_r of user r is a concave function of $\mathbf{w}_r$ and a continuous function of $\mathbf{w}$. To prove this, it suffices to show that $U_r(d_r(\mathbf{x}_r(\mathbf{w}_r; \mathbf{w}_{-r})))$ is a concave function of $\mathbf{w}_r$ and a continuous function of $\mathbf{w}$. We first observe that by Proposition 3.12, $x_{jr}(\mathbf{w})$ is a concave function of $w_{jr} \geq 0$, and a continuous function of $\mathbf{w}$. Since for each j the function $x_{jr}(\mathbf{w})$ does not depend on w_{kr} for $k \neq j$, we conclude that each component of $\mathbf{x}_r(\mathbf{w}_r; \mathbf{w}_{-r})$ is a concave function of $\mathbf{w}_r$. Now since d_r defined as the optimal objective value of a linear program, $d_r(\overline{\mathbf{x}}_r)$ is continuous and concave as a function of $\overline{\mathbf{x}}_r$ [15]. In addition, $d_r(\overline{\mathbf{x}}_r)$ is *nondecreasing* in $\overline{\mathbf{x}}_r$; i.e., if $\overline{x}_{jr} \geq \hat{x}_{jr}$ for all j, then $d_r(\overline{\mathbf{x}}_r) \geq d_r(\hat{\mathbf{x}}_r)$ (this follows from the problem (3.68)-(3.70)). These properties of x_{jr} and d_r, combined with the fact that U_r is concave, continuous, and nondecreasing from Assumption 3.5, imply that $U_r(d_r(\mathbf{x}_r(\mathbf{w}_r; \mathbf{w}_{-r})))$ is a concave function of $\mathbf{w}_r$, and a continuous function of $\mathbf{w}$. □

The following theorem demonstrates that the utility lost at any Nash equilibrium is no worse than $4\sqrt{2} - 5$ of the maximum possible aggregate surplus, matching the result derived for the single link model. The proof follows the proof of Theorem 2.13: we construct a single link game at each link j, whose Nash equilibrium is the same as the fixed Nash equilibrium of the network game. We then apply Theorem 3.8 at each link to complete the proof. However, we note that this result does not require U_r to be strictly increasing or continuously differentiable, and is therefore a stronger version of Theorem 3.8 for the single link case.

Theorem 3.14

Suppose that Assumptions 3.5-3.7 hold. Assume also that $U_r(0) \geq 0$ for all users r. If $\mathbf{w}$ is any Nash equilibrium for the game defined by $(T_1, \ldots, T_R)$, and $(\mathbf{y}^S, \mathbf{f}^S, \mathbf{d}^S)$ is any optimal solution to SYSTEM, then:

$$\sum_r U_r(d_r(\mathbf{x}_r(\mathbf{w}))) - \sum_j C_j(f_j(\mathbf{w})) \geq (4\sqrt{2} - 5) \left(\sum_r U_r(d_r^S) - \sum_j C_j(f_j^S) \right).$$

Proof. The proof consists of three main steps. First, we describe the entire problem in terms of the vector $\mathbf{x}_r(\mathbf{w}) = (x_{jr}(\mathbf{w}), j \in J)$ of the rate allocations to user r from the network. We show in Lemma 3.15 that Nash equilibria can be characterized in terms of each user r optimally choosing a rate allocation $\overline{\mathbf{x}}_r = (\overline{x}_{jr}, j \in J)$, given the vector of bids $\mathbf{w}_{-r}$ of all other users.

In the second step, we observe that the utility to user r given a vector of rate allocations $\overline{\mathbf{x}}_r$ is exactly $U_r(d_r(\overline{\mathbf{x}}_r))$; we call this a "composite" utility function. In Lemma 3.16, we linearize this composite utility function. Formally, we replace $U_r(d_r(\overline{\mathbf{x}}_r))$ with a linear function $\boldsymbol{\alpha}_r^\top \overline{\mathbf{x}}_r$. The difficulty in this phase of the analysis is that the composite utility function $U_r(d_r(\cdot))$ may not be differentiable, because the max-flow function $d_r(\cdot)$ is not differentiable everywhere; as a result, convex analytic techniques are required.

Finally, we conclude the proof by observing that when the "composite" utility function for user r is linear in the vector of rate allocations $\overline{\mathbf{x}}_r$, the network structure is no longer relevant. In this case the game defined by $(Q_1, \ldots, Q_R)$ decouples into J games, one for each link. We then apply Theorem 3.8 at each link to arrive at the bound in the theorem.

We start by describing the problem in terms of the vector $\mathbf{x}_r(\mathbf{w}) = (x_{jr}(\mathbf{w}), j \in J)$ of the rate allocations to user r from the network. We redefine the problem *SYSTEM* as follows:

$$\text{maximize} \quad \sum_r U_r(d_r(\overline{\mathbf{x}}_r)) - \sum_j C_j(f_j) \tag{3.73}$$

$$\text{subject to} \quad \sum_r \overline{x}_{jr} = f_j, \quad j \in J; \tag{3.74}$$

$$\overline{x}_{jr} \geq 0, \quad j \in J, \ r \in R. \tag{3.75}$$

(The notation $\overline{\mathbf{x}}_r$ is used here to distinguish from the function $\mathbf{x}_r(\mathbf{w})$.) In this problem, the network only chooses how to allocate rate at each link to the users. The users then solve a max-flow problem to determine the maximum rate at which they can send (this is captured by the function $d_r(\cdot)$). This problem is equivalent to the problem *SYSTEM* as defined in (3.62)-(3.65), because of the definition of $d_r(\cdot)$ in (3.68)-(3.70). We label an optimal solution to this problem by $(\mathbf{x}_r^S, r \in R; f_j^S, j \in J)$.

Our next step is to show that a Nash equilibrium may be characterized in terms of users optimally choosing rate allocations $(\overline{\mathbf{x}}_r, r \in R)$. We begin by "inverting" the function $x_{jr}(\mathbf{w})$, with respect to w_{jr}; that is, we determine the amount that user r must pay to link j to receive a predetermined rate allocation $\overline{x}_{jr}$, given that all other users have bid $\mathbf{w}_{-r}$. Formally, we observe from Proposition 3.12 that $x_{jr}(\mathbf{w})$ is concave, strictly increasing, and continuous in w_{jr}. Finally, since $x_{jr}(\mathbf{w}) = 0$ if $w_{jr} = 0$, and $x_{jr}(\mathbf{w}) \to \infty$ as $w_{jr} \to \infty$, we can define a function $\omega_{jr}(\overline{x}_{jr}; \mathbf{w}_{-r})$ for $\overline{x}_{jr} \geq 0$, which satisfies:

$$x_{jr}(\mathbf{w}) = \overline{x}_{jr} \quad \text{if and only if} \quad w_{jr} = \omega_{jr}(\overline{x}_{jr}; \mathbf{w}_{-r}).$$

From the properties of x_{jr} described above, we note that for a fixed vector $\mathbf{w}_{-r}$, the function $\omega_{jr}(\cdot; \mathbf{w}_{-r})$ is convex, strictly increasing, and continuous, with $\omega_{jr}(0; \mathbf{w}_{-r}) = 0$ and $\omega_{jr}(\overline{x}_{jr}; \mathbf{w}_{-r}) \to \infty$ as $\overline{x}_{jr} \to \infty$.

We now use the functions ω_{jr} to write user r's payoff in terms of the allocated rate vector $\overline{\mathbf{x}}_r = (\overline{x}_{jr}, j \in J)$, rather than in terms of the bid $\mathbf{w}_r$. For $\overline{\mathbf{x}}_r \geq 0$, we define a function $F_r(\overline{\mathbf{x}}_r; \mathbf{w}_{-r})$ as follows:

$$F_r(\overline{\mathbf{x}}_r; \mathbf{w}_{-r}) = U_r(d_r(\overline{\mathbf{x}}_r)) - \sum_j \omega_{jr}(\overline{x}_{jr}; \mathbf{w}_{-r}). \tag{3.76}$$

We now have the following lemma, which shows a Nash equilibrium may be characterized by an optimal choice of $\overline{\mathbf{x}}_r$ for each r.

Lemma 3.15 *A vector $\mathbf{w}$ is a Nash equilibrium if and only if the following condition holds for each user r:*

$$\mathbf{x}_r(\mathbf{w}) \in \arg\max_{\overline{\mathbf{x}}_r \geq 0} F_r(\overline{\mathbf{x}}_r; \mathbf{w}_{-r}). \tag{3.77}$$

Proof of Lemma. Fix a bid vector $\mathbf{w}$, and suppose that there exists a vector $\overline{\mathbf{x}}_r \geq 0$ such that:

$$F_r(\overline{\mathbf{x}}_r; \mathbf{w}_{-r}) > F_r(\mathbf{x}_r(\mathbf{w}); \mathbf{w}_{-r}). \tag{3.78}$$

Since $\omega_{jr}(x_{jr}(\mathbf{w}); \mathbf{w}_{-r}) = w_{jr}$, we have $F_r(\mathbf{x}_r(\mathbf{w}); \mathbf{w}_{-r}) = T_r(\mathbf{w}_r; \mathbf{w}_{-r})$. Now consider the bid vector $\overline{\mathbf{w}}_r$ defined by $\overline{w}_{jr} = \omega_{jr}(\overline{x}_{jr}; \mathbf{w}_{-r})$. Then $x_{jr}(\overline{\mathbf{w}}_r; \mathbf{w}_{-r}) = \overline{x}_{jr}$ for each j, so:

$$T_r(\overline{\mathbf{w}}_r; \mathbf{w}_{-r}) = F_r(\overline{\mathbf{x}}_r; \mathbf{w}_{-r}).$$

Thus $\overline{\mathbf{w}}_r$ is a profitable deviation for user r, so $\mathbf{w}$ could not have been a Nash equilibrium.

Conversely, suppose that $\mathbf{w}$ is not a Nash equilibrium. As above, we have the equality $F_r(\mathbf{x}_r(\mathbf{w}); \mathbf{w}_{-r}) = T_r(\mathbf{w}_r; \mathbf{w}_{-r})$. Fix a user r, and let $\overline{\mathbf{w}}_r$ be a profitable deviation for user r, so that $T_r(\overline{\mathbf{w}}_r; \mathbf{w}_{-r}) > T_r(\mathbf{w}_r; \mathbf{w}_{-r})$. For each j, let $\overline{x}_{jr} = x_{jr}(\overline{\mathbf{w}}_r; \mathbf{w}_{-r})$. Then $\omega_{jr}(\overline{x}_{jr}; \mathbf{w}_{-r}) = w_{jr}$, so that $F_r(\overline{\mathbf{x}}_r; \mathbf{w}_{-r}) = T_r(\overline{\mathbf{w}}_r; \mathbf{w}_{-r})$. Thus we conclude that $F_r(\overline{\mathbf{x}}_r; \mathbf{w}_{-r}) > F_r(\mathbf{x}_r(\mathbf{w}); \mathbf{w}_{-r})$, so that (3.77) does not hold. □

Now suppose that $\mathbf{w}$ is a Nash equilibrium. Our approach is to replace user r by J users (which we call "virtual" users), one at each link j; this process has the effect of *isolating* each of the links, and removes any dependence on network structure. We define the virtual users so that $\mathbf{w}$ remains a Nash equilibrium at each single link game. Formally, for each user r, we construct a vector $\boldsymbol{\alpha}_r = (\alpha_{jr}, j \in J)$, and consider a single link game at each link j where user r has linear utility function $U_{jr}(x_{jr}) = \alpha_{jr} x_{jr}$. We choose the vectors $\boldsymbol{\alpha}_r$ so that the Nash equilibrium at each single link game is also given by $\mathbf{w}$; we then apply the result of Theorem 3.8 for the single link model to complete the proof of the theorem.

As in the proof of Theorem 2.13, a technical difficulty arises here because the function $U_r(d_r(\cdot))$ may not be differentiable. If the composite function $g_r = U_r(d_r(\cdot))$ were differentiable, then as in the proof of Theorem 3.8, we could find an appropriate vector $\boldsymbol{\alpha}_r$ by choosing $\boldsymbol{\alpha}_r = \nabla g_r(\mathbf{x}_r(\mathbf{w}))$. However, in general $U_r(d_r(\cdot))$ is not differentiable; instead, we will choose $\boldsymbol{\alpha}_r$ to be a *supergradient* of $U_r(d_r(\cdot))$, i.e., we require $-\boldsymbol{\alpha}_r$ to be a *subgradient* of $-U_r(d_r(\cdot))$. The reader is referred to the Notation section for reference on these definitions from convex analysis. The key relationship we note is that $\boldsymbol{\gamma}$ is a supergradient of an extended real-valued function $g : \mathbb{R}^J \to \mathbb{R}$ at $\mathbf{x}$ if and only if for all $\overline{\mathbf{x}} \in \mathbb{R}^J$:

$$g(\overline{\mathbf{x}}) \leq g(\mathbf{x}) + \boldsymbol{\gamma}^\top (\overline{\mathbf{x}} - \mathbf{x}).$$

Lemma 3.15 allows us to characterize the Nash equilibrium $\mathbf{w}$ as a choice of optimal rate allocation $\overline{\mathbf{x}}_r$ by each user r, given the strategy vector $\mathbf{w}_{-r}$ of all other users. We recall the definition of F_r in (3.76); we will now view F_r as an extended real valued function, by defining $F_r(\overline{\mathbf{x}}_r; \mathbf{w}_{-r}) = -\infty$ if $\overline{x}_{jr} < 0$ for some j. We also define extended real-valued functions G_r and K_{jr} on $\mathbb{R}^J$ as follows:

$$G_r(\overline{\mathbf{x}}_r) = \begin{cases} U_r(d_r(\overline{\mathbf{x}}_r)), & \text{if } \overline{\mathbf{x}}_r \geq 0; \\ -\infty, & \text{otherwise.} \end{cases}$$

and

$$K_{jr}(\overline{\mathbf{x}}_r; \mathbf{w}_{-r}) = \begin{cases} -\omega_{jr}(\overline{x}_{jr}; \mathbf{w}_{-r}), & \text{if } \overline{x}_{jr} \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

Then we have $F_r = G_r + \sum_j K_{jr}$ on $\mathbb{R}^J$. The following lemma establishes existence of the desired vector $\boldsymbol{\alpha}_r$.

Lemma 3.16 *Let $\mathbf{w}$ be a Nash equilibrium. Then for each user r, there exists a vector $\boldsymbol{\alpha}_r = (\alpha_{jr}, j \in J) \geq 0$ such that $\boldsymbol{\alpha}_r \in -\partial[-G_r(\mathbf{x}_r(\mathbf{w}))]$, and the following relation holds:*

$$\mathbf{x}_r(\mathbf{w}) \in \arg\max_{\overline{\mathbf{x}}_r \geq 0} \left[\boldsymbol{\alpha}_r^\top \overline{\mathbf{x}}_r - \sum_j \omega_{jr}(\overline{x}_{jr}; \mathbf{w}_{-r}) \right]. \qquad (3.79)$$

Proof of Lemma. Fix a user r. We observe that G_r is a concave function of $\overline{\mathbf{x}}_r \in \mathbb{R}^J$. This follows as in the proof of Proposition 3.13, because d_r is a concave function of its argument (as it is the optimal objective value of the linear program (3.68)-(3.70)), and U_r is nondecreasing and concave. Furthermore, we note that $K_{jr}(\overline{\mathbf{x}}_r; \mathbf{w}_{-r})$ is a concave function of $\overline{\mathbf{x}}_r \in \mathbb{R}^J$ as well, since $\omega_{jr}(\overline{x}_{jr}, \mathbf{w}_{-r})$ is convex and nonnegative for $\overline{x}_{jr} \geq 0$. Consequently, F_r is a concave function of $\overline{\mathbf{x}}_r \in \mathbb{R}^J$. In particular, $-F_r$, $-G_r$, and $-K_{jr}$ are proper, convex, extended real-valued functions. It is straightforward to show, using Theorem 23.8 in [103], that at $\mathbf{x}_r(\mathbf{w})$ we have:

$$\partial[-F_r(\mathbf{x}_r(\mathbf{w}); \mathbf{w}_{-r})] = \partial[-G_r(\mathbf{x}_r(\mathbf{w}))] + \sum_j \partial[-K_{jr}(\mathbf{x}_r(\mathbf{w}); \mathbf{w}_{-r})]. \tag{3.80}$$

(The summation here of the subdifferentials on the right hand side is a summation of sets, where $A + B = \{\mathbf{x} + \mathbf{y} : \mathbf{x} \in A, \mathbf{y} \in B\}$; if either A or B is empty, then $A + B$ is empty as well.)

Since $\mathbf{w}$ is a Nash equilibrium, from Lemma 3.15, we have for all $\overline{\mathbf{x}}_r \geq 0$ that:

$$F_r(\mathbf{x}_r(\mathbf{w}); \mathbf{w}_{-r}) \geq F_r(\overline{\mathbf{x}}_r; \mathbf{w}_{-r}).$$

Since $F_r(\overline{\mathbf{x}}_r; \mathbf{w}_{-r}) = -\infty$ if there exists a j such that $\overline{x}_{jr} < 0$, we must in fact have $F_r(\mathbf{x}_r(\mathbf{w}); \mathbf{w}_{-r}) \geq F_r(\overline{\mathbf{x}}_r; \mathbf{w}_{-r})$ for *all* $\overline{\mathbf{x}}_r \in \mathbb{R}^J$ so we conclude $\mathbf{0}$ is a supergradient of $-F_r(\cdot; \mathbf{w}_{-r})$ at $\mathbf{x}_r(\mathbf{w})$. As a result, it follows from (3.80) that there exist vectors $\boldsymbol{\alpha}_r$ and $\boldsymbol{\beta}_{jr}$ with $\boldsymbol{\alpha}_r \in -\partial[-G_r(\mathbf{x}_r(\mathbf{w}))]$ and $\boldsymbol{\beta}_{jr} \in -\partial[-K_{jr}(\mathbf{x}_r(\mathbf{w}); \mathbf{w}_{-r})]$, such that $\boldsymbol{\alpha}_r = -\sum_j \boldsymbol{\beta}_{jr}$.

We first note that $G_r(\overline{\mathbf{x}}_r)$ is a nondecreasing function of $\overline{\mathbf{x}}_r$; that is, if $\overline{\mathbf{x}}_r \geq \hat{\mathbf{x}}_r$, then $G_r(\overline{\mathbf{x}}_r) \geq G_r(\hat{\mathbf{x}}_r)$. From this fact it follows that $\boldsymbol{\alpha}_r$ must be nonnegative, i.e., $\alpha_{jr} \geq 0$ for all j. It remains to be shown that (3.79) holds. We observe that $\mathbf{0}$ is a supergradient of the following function at $\mathbf{x}_r(\mathbf{w})$:

$$\hat{F}_r(\overline{\mathbf{x}}_r; \mathbf{w}_{-r}) = \begin{cases} \boldsymbol{\alpha}_r^\top \overline{\mathbf{x}}_r - \sum_j \omega_{jr}(\overline{x}_{jr}; \mathbf{w}_{-r}), & \text{if } \overline{\mathbf{x}}_r \geq 0; \\ -\infty, & \text{otherwise.} \end{cases}$$

This observation follows by replacing $G_r(\overline{\mathbf{x}}_r)$ with the following function $\hat{G}_r$ on $\mathbb{R}^J$:

$$\hat{G}_r(\overline{\mathbf{x}}_r) = \begin{cases} \boldsymbol{\alpha}_r^\top \overline{\mathbf{x}}_r, & \text{if } \overline{\mathbf{x}}_r \geq 0; \\ -\infty, & \text{otherwise.} \end{cases}$$

Then we have $\hat{F}_r = \hat{G}_r + \sum_j K_{jr}$; and as before:

$$\partial[-\hat{F}_r(\mathbf{x}_r(\mathbf{w}); \mathbf{w}_{-r}))] = \partial[-\hat{G}_r(\mathbf{x}_r(\mathbf{w}))] + \sum_j \partial[-K_{jr}(\mathbf{x}_r(\mathbf{w}); \mathbf{w}_{-r})].$$

The vector $-\boldsymbol{\alpha}_r$ is a subgradient of $-\hat{G}_r$ for all $\overline{\mathbf{x}}_r \geq 0$. In particular, we can conclude that $\boldsymbol{\alpha}_r \in -\partial[-\hat{G}_r(\mathbf{x}_r(\mathbf{w}))]$. Recall that we have already shown $\boldsymbol{\alpha}_r = -\sum_j \boldsymbol{\beta}_{jr} \in \sum_j -\partial[-K_{jr}(\mathbf{x}_r(\mathbf{w}); \mathbf{w}_{-r})]$. Thus $\mathbf{0} \in \partial[-\hat{F}_r(\mathbf{x}_r(\mathbf{w}); \mathbf{w}_{-r})]$. This implies (3.79), as required. □

Let $\mathbf{w}$ be a Nash equilibrium. For each user r, fix the vector $\boldsymbol{\alpha}_r$ given by the preceding lemma. We start by observing that for each user r, since $\boldsymbol{\alpha}_r$ is a supergradient of $G_r(\mathbf{x}_r(\mathbf{w}))$, we have:

$$U_r(d_r(\mathbf{x}_r^S)) \leq U_r(d_r(\mathbf{x}_r(\mathbf{w}))) + \boldsymbol{\alpha}_r^\top(\mathbf{x}_r^S - \mathbf{x}_r(\mathbf{w})). \tag{3.81}$$

Now note that since ω_{jr} is strictly increasing, if $\boldsymbol{\alpha}_r = \mathbf{0}$, then the unique maximizer in (3.79) is $\overline{\mathbf{x}}_r = \mathbf{0}$. Thus if $\boldsymbol{\alpha}_r = \mathbf{0}$ for all r, we must have $\mathbf{x}_r(\mathbf{w}) = 0$ for all r. But from (3.81), we have the following trivial inequality:

$$\sum_r U_r(d_r(\mathbf{x}_r(\mathbf{w}))) \geq \sum_r U_r(d_r(\mathbf{x}_r^S)).$$

Since $d_r(\mathbf{x}_r(\mathbf{w})) = d_r(\mathbf{0}) = 0$ for all r, this is only possible if $U_r(d_r(\mathbf{x}_r^S)) = U_r(0)$ for all r as well. It follows that the aggregate surplus is zero at both the Nash equilibrium and the optimal solution to *SYSTEM*, so the theorem holds in this case. We may assume without loss of generality, therefore, that $\boldsymbol{\alpha}_r \neq 0$ for at least one user r.

We have the following simplification of (3.79):

$$\begin{aligned}
\mathbf{x}_r(\mathbf{w}) &\in \arg\max_{\overline{\mathbf{x}}_r \geq 0} \left[\boldsymbol{\alpha}_r^\top \overline{\mathbf{x}}_r - \sum_j \omega_{jr}(\overline{x}_{jr}; \mathbf{w}_{-r}) \right] \\
&= \arg\max_{\overline{\mathbf{x}}_r \geq 0} \left[\sum_j \left(\alpha_{jr}\overline{x}_{jr} - \omega_{jr}(\overline{x}_{jr}; \mathbf{w}_{-r}) \right) \right].
\end{aligned}$$

The maximum on the right hand side of the preceding expression decomposes into separate maximizations for each link j. We conclude that for each link j, we have in fact:

$$x_{jr}(\mathbf{w}) \in \arg\max_{\overline{x}_{jr} \geq 0} \left[\alpha_{jr}\overline{x}_{jr} - \omega_{jr}(\overline{x}_{jr}; \mathbf{w}_{-r}) \right]. \tag{3.82}$$

Fix a link j. We view the users as playing a single link game at link j, with utility function for user r given by $U_{jr}(x_{jr}) = \alpha_{jr}x_{jr}$. The preceding expression states that (3.77) in Lemma 3.15 is satisfied, so we conclude that $\mathbf{w}$ is a Nash equilibrium for this single link game at link j. More precisely, we have that $(w_{j1}, \ldots, w_{jR})$ is a Nash equilibrium for the single link game at link j, when R users with utility functions $(U_{j1}, \ldots, U_{jR})$ compete for link j. The maximum aggregate surplus for this link is given by:

$$\max_{\overline{x}_{j1},\ldots,\overline{x}_{jR}\geq 0}\left[\sum_r \alpha_{jr}\overline{x}_{jr} - C_j\left(\sum_r \overline{x}_{jr}\right)\right] = \max_{\overline{f}_j\geq 0}\left[(\max_r \alpha_{jr})\overline{f}_j - C_j(\overline{f}_j)\right].$$

Now if $\alpha_{jr} = 0$, then from (3.82), the optimal choice for user r is $x_{jr}(\mathbf{w}) = 0$. Thus there are two possibilities: either $\alpha_{jr} = 0$ for all r, in which case both the Nash equilibrium aggregate surplus and maximum aggregate surplus are zero; or $\alpha_{jr} > 0$ for at least one user r, in which case the maximum aggregate surplus is strictly positive, and we can apply Theorem 3.8 to find:

$$\sum_r \alpha_{jr}x_{jr}(\mathbf{w}) - C_j(f_j(\mathbf{w})) \geq \left(4\sqrt{2}-5\right)\left(\max_{\overline{f}_j\geq 0}\left[(\max_r \alpha_{jr})\overline{f}_j - C_j(\overline{f}_j)\right]\right). \tag{3.83}$$

In particular, note that the preceding inequality holds for all links j (since it holds trivially for those links where $\alpha_{jr} = 0$ for all r).

We now complete the proof of the theorem, by following the proof of Step 1 of Theorem 3.8. We first note that we can assume without loss of generality that $\sum_r U_r(d_r^S) - \sum_j C_j(f_j^S) > 0$. If not, then $\sum_r U_r(d_r^S) - \sum_j C_j(f_j^S) = 0$. On the other hand, since $\mathbf{w}$ is a Nash equilibrium, for each r we have:

$$U_r(d_r(\mathbf{x}_r(\mathbf{w}))) - \sum_j w_{jr} \geq 0.$$

This holds since user r can guarantee a payoff of zero by bidding $\mathbf{w}_r = \mathbf{0}$. Summing over all users, and applying (3.66), we have:

$$\sum_r U_r(d_r(\mathbf{x}_r(\mathbf{w}))) - \sum_j f_j(\mathbf{w})p_j(f_j(\mathbf{w})) \geq 0.$$

By convexity, we know $f_j(\mathbf{w})p_j(f_j(\mathbf{w})) \geq C_j(f_j(\mathbf{w}))$; thus:

$$\sum_r U_r(d_r(\mathbf{x}_r(\mathbf{w}))) - \sum_j C_j(f_j(\mathbf{w})) \geq 0.$$

Thus if $\sum_r U_r(d_r^S) - \sum_j C_j(f_j^S) = 0$, it must also be the case that $\sum_r U_r(d_r(\mathbf{x}_r(\mathbf{w}))) - \sum_j C_j(f_j(\mathbf{w})) = 0$, since $(\mathbf{y}^S, \mathbf{f}^S, \mathbf{d}^S)$ is an optimal solution to *SYSTEM*. We conclude the theorem trivially holds in this case, so we can assume without loss of generality that $\sum_r U_r(d_r^S) - \sum_j C_j(f_j^S) > 0$.

We note that we have:

$$\sum_r \boldsymbol{\alpha}_r^\top \mathbf{x}_r^S - \sum_j C_j(f_j^S) = \sum_j \left(\sum_r \alpha_{jr} x_{jr}^S - C_j(f_j^S) \right) \leq \sum_j \max_{\overline{f}_j \geq 0} \left[(\max_r \alpha_{jr}) \overline{f}_j - C_j(\overline{f}_j) \right]. \tag{3.84}$$

We now reason as follows, using (3.81) for the first inequality, and (3.84) for the second:

$$\frac{\sum_r U_r(d_r(\mathbf{x}_r(\mathbf{w}))) - \sum_j C_j(f_j(\mathbf{w}))}{\sum_r U_r(d_r(\mathbf{x}_r^S)) - \sum_j C_j(f_j^S)} \tag{3.85}$$

$$\geq \frac{\sum_r \big(U_r(d_r(\mathbf{x}_r(\mathbf{w}))) - \boldsymbol{\alpha}_r^\top \mathbf{x}_r(\mathbf{w})\big) + \sum_r \boldsymbol{\alpha}_r^\top \mathbf{x}_r(\mathbf{w}) - \sum_j C_j(f_j(\mathbf{w}))}{\sum_r \left(U_r(d_r(\mathbf{x}_r(\mathbf{w}))) + \boldsymbol{\alpha}_r^\top (\mathbf{x}_r^S - \mathbf{x}_r(\mathbf{w}))\right) - \sum_j C_j(f_j^S)}$$

$$= \frac{\sum_r \big(U_r(d_r(\mathbf{x}_r(\mathbf{w}))) - \boldsymbol{\alpha}_r^\top \mathbf{x}_r(\mathbf{w})\big) + \sum_r \boldsymbol{\alpha}_r^\top \mathbf{x}_r(\mathbf{w}) - \sum_j C_j(f_j(\mathbf{w}))}{\sum_r \left(U_r(d_r(\mathbf{x}_r(\mathbf{w}))) - \boldsymbol{\alpha}_r^\top \mathbf{x}_r(\mathbf{w})\right) + \sum_r \boldsymbol{\alpha}_r^\top \mathbf{x}_r^S - \sum_j C_j(f_j^S)}$$

$$\geq \frac{\sum_r \big(U_r(d_r(\mathbf{x}_r(\mathbf{w}))) - \boldsymbol{\alpha}_r^\top \mathbf{x}_r(\mathbf{w})\big) + \sum_j \left(\sum_r \alpha_{jr} x_{jr}(\mathbf{w}) - C_j(f_j(\mathbf{w}))\right)}{\sum_r \left(U_r(d_r(\mathbf{x}_r(\mathbf{w}))) - \boldsymbol{\alpha}_r^\top \mathbf{x}_r(\mathbf{w})\right) + \sum_j \max_{\overline{f}_j \geq 0} \left[(\max_r \alpha_{jr})\overline{f}_j - C_j(\overline{f}_j)\right]}. \tag{3.86}$$

Since $U_r(d_r(\mathbf{0})) = U_r(0) \geq 0$, by concavity of U_r and the fact that $\boldsymbol{\alpha}_r \in \partial[-G_r(\mathbf{x}_r(\mathbf{w}))]$ we have:

$$U_r(d_r(\mathbf{x}_r(\mathbf{w}))) - \boldsymbol{\alpha}_r^\top \mathbf{x}_r(\mathbf{w}) \geq 0.$$

Furthermore, from (3.20), we have $\alpha_{jr} > p_j(f_j(\mathbf{w}))$ if $x_{jr}(\mathbf{w}) > 0$; thus $\sum_r \alpha_{jr} x_{jr}(\mathbf{w}) > f_j(\mathbf{w}) p_j(f_j(\mathbf{w})) \geq C_j(f_j(\mathbf{w}))$, where the second inequality follows by convexity (Assumption 3.6). This yields:

$$0 < \sum_j \left(\sum_r \alpha_{jr} x_{jr}(\mathbf{w}) - C_j(f_j(\mathbf{w})) \right) \leq \sum_j \max_{\overline{f}_j \geq 0} \left[(\max_r \alpha_{jr}) \overline{f}_j - C_j(\overline{f}_j) \right].$$

So we conclude from relations (3.83) and (3.86) that:

$$\frac{\sum_r U_r(d_r(\mathbf{x}_r(\mathbf{w}))) - \sum_j C_j(f_j(\mathbf{w}))}{\sum_r U_r(d_r(\mathbf{x}_r^S)) - \sum_j C_j(f_j^S)} \geq \frac{\sum_j \left(\sum_r \alpha_{jr} x_{jr}(\mathbf{w}) - C_j(f_j(\mathbf{w}))\right)}{\sum_j \max_{\overline{f}_j \geq 0} \left[(\max_r \alpha_{jr})\overline{f}_j - C_j(\overline{f}_j)\right]} \geq 4\sqrt{2} - 5.$$

(Observe that both denominators in this chain of inequalities are nonzero.) Since $\mathbf{w}$ was assumed to be a Nash equilibrium, this completes the proof of the theorem. □

The preceding theorem uses the bound on efficiency loss in the single link game to bound the efficiency loss when users are price anticipating in general networks. Note that since we knew from Theorem 3.8 that the bound of $4\sqrt{2}-5$ was essentially tight for single link games, and a single link is a special case of a general network, the bound $4\sqrt{2}-5$ is also tight in this setting.

We conclude with two observations. First, the discussion of Section 2.4.3 remains relevant here. The network pricing mechanism discussed by Kelly et al. in [65] involves each user submitting only their *total payment* to the network; the network then calculates a rate allocation. Following an argument similar to Proposition 2.17, it is possible to show that at a Nash equilibrium of the game considered in this section, the resulting allocation for users with positive bids is identical to that obtained if users submit only their total payments $(w_1, \ldots, w_R)$ (where $w_r = \sum_j w_{jr}$) to the mechanism proposed by Kelly et al. in [65]. However, an open question remains concerning the existence and efficiency of Nash equilibria of the game where users submit only their total payments to the network. Recall from the discussion in Section 2.4.3 that Hajek and Yang have shown the efficiency loss can be arbitrarily high when users are price anticipating in the game where they submit only their total payments, and link capacities are inelastic [53]. We thus conjecture that in the worst case, efficiency loss may be arbitrarily high even when link capacities are elastic, based on the intuition that the model considered by Hajek and Yang can be viewed as a limit of a sequence of games with elastic link capacities.

Our second observation is that, as in Section 2.5.2, the essential structure in the network game we consider here is that the function $U_r(d_r(\mathbf{x}_r))$ is a concave and continuous function of the vector $\mathbf{x}_r \geq 0$, and also *nondecreasing*; that is, if $x_{jr} \geq \overline{x}_{jr}$ for all $j \in J$, then $U_r(d_r(\mathbf{x}_r)) \geq U_r(d_r(\overline{\mathbf{x}}_r))$. Thus, arguing exactly as in Section 2.5.2, we can consider a more general resource allocation game where the utility to user r is a concave, continuous, nondecreasing function of the vector of resources allocated, $V_r(\mathbf{x}_r)$; all the results of this section continue to hold for this more general game.

■ 3.5 Cournot Competition

In this section, we will consider a game where the strategies of users are their desired rates, rather than their total payments; such games, where the strategy of the market participants is the quantity demanded or supplied, are known as *Cournot games* [82, 134]. Such games are not well defined when supply is inelastic, since in that case a market mechanism has no means to ensure supply equals demand if the sum of the desired rates of the users exceeds supply. On the other hand, when supply is elastic, then the Cournot mechanism is valid; we will find, however, that in general a Cournot mechanism can have arbitrarily high efficiency loss.

Formally, we consider the following model. As before, R users share a single communication link. We continue to assume that each user r has a utility function U_r, and that the link has a cost function C. We make the following assumptions.

Assumption 3.8
For each r, over the domain $d_r \geq 0$ the utility function $U_r(d_r)$ is concave, nondecreasing, and continuous; and over the domain $d_r > 0$, $U_r(d_r)$ is continuously differentiable. Furthermore, the right directional derivative at 0, denoted $U_r'(0)$, is finite.

Assumption 3.9
There exists a continuous, convex, nondecreasing function $p(f)$ over $f \geq 0$ with $p(0) \geq 0$ and $p(f) \to \infty$ as $f \to \infty$, such that for $f \geq 0$:

$$C(f) = \int_0^f p(z)dz.$$

Thus $C(f)$ is convex and nondecreasing.

Assumption 3.8 is nearly identical to Assumption 3.5, though we also impose the requirement that U_r is differentiable for ease of technical presentation. Note that Assumption 3.9 is more general than Assumption 3.2; for example, it allows $p(0)$ to be nonzero. Before continuing, therefore, we consider the implications of Assumption 3.9 on the model of Section 3.1, and in particular on the conclusion of Theorem 3.8. Unfortunately, we show in the next example that there exist price functions p satisfying Assumption 3.9 for which the efficiency loss may be arbitrarily high when users are price anticipating.

Example 3.3
Fix $a < 1$, and consider a system consisting of $R = 2$ users, with $U_1(d_1) = d_1$, and $U_2(d_2) = ad_2$. Define the price function p according to $p(f) = a + (f-1)^+$ (i.e., $p(f) = a$ if $f \leq 1$, and $p(f) = a + f - 1$ if $f \geq 1$). Let $C(f) = \int_0^f p(z)\ dz$ be the associated cost function. Then it is easy to verify that p and C satisfy Assumption 3.9. Furthermore, for this price function it can be shown that a vector $\mathbf{w}$ is a Nash equilibrium of the game defined by $(Q_1, \ldots, Q_R)$ (where Q_r is defined in (3.11)) if and only if $\mathbf{w}$ satisfies the conditions of Proposition 3.6. Define the vector $\mathbf{w}$ according to $w_1 = a - a^3$, and $w_2 = a^3$; we claim that $\mathbf{w}$ is a Nash equilibrium. First note that there holds $f(\mathbf{w}) = 1$, $p(f(\mathbf{w})) = a$, $\beta^+(f(\mathbf{w})) = 1/(1+a)$, and $\beta^-(f(\mathbf{w})) = 0$, with $d_1(\mathbf{w}) = 1 - a^2$, and $d_2(\mathbf{w}) = a^2$. Recall the Nash equilibrium conditions (3.19)-(3.20) for each r:

$$U_r'(d_r(\mathbf{w}))\left(1 - \beta^+(f(\mathbf{w})) \cdot \frac{d_r(\mathbf{w})}{f(\mathbf{w})}\right) \leq p(f(\mathbf{w}));$$
$$U_r'(d_r(\mathbf{w}))\left(1 - \beta^-(f(\mathbf{w})) \cdot \frac{d_r(\mathbf{w})}{f(\mathbf{w})}\right) \geq p(f(\mathbf{w})), \quad \text{if } w_r > 0.$$

Using these conditions, it follows that $\mathbf{w}$ is a Nash equilibrium. Furthermore, it can be verified that the Nash equilibrium aggregate surplus in this case is given by $1 - a^2 + a^3 - a = (1-a)^2(1+a)$. On the other hand, the maximal aggregate surplus is achieved when $p(f^S) = 1$, i.e., when $f^S = 2 - a$; and in this case the maximal aggregate surplus is $1 - a + (1-a)^2/2$. Thus the ratio of Nash equilibrium aggregate surplus to maximal aggregate surplus is:

$$\frac{(1-a)^2(1+a)}{1 - a + \frac{1}{2}(1-a)^2} = \frac{(1-a)(1+a)}{1 + \frac{1}{2}(1-a)}.$$

As $a \to 1$, the right hand side approaches zero; thus the efficiency loss may be arbitrarily high at a Nash equilibrium. □

Because of the negative result of the previous example, we focused only on price functions p satisfying Assumption 3.2 for the duration of Sections 3.1 and 3.2. For the development of the current section, however, the arguments are in fact simplified if we only require the price function p to satisfy the more general Assumption 3.9.

We continue to assume the problem *SYSTEM* is defined as in (3.1)-(3.2):

SYSTEM:

$$\begin{aligned} \text{maximize} \quad & \sum_r U_r(d_r) - C\left(\sum_r d_r\right) \\ \text{subject to} \quad & d_r \geq 0, \quad r = 1, \ldots, R. \end{aligned}$$

As in Section 3.1, the objective function of this problem is the *aggregate surplus* (see Section 1.1). Since $p(f) \to \infty$ as $f \to \infty$, while U_r only grows at most linearly, it follows that an optimal solution exists. We now consider the following pricing scheme for rate allocation. Each user r chooses a desired rate d_r; the network manager then sets a single price $\mu(\mathbf{d})$. In this case, given a price $\mu > 0$, user r chooses d_r to maximize:

$$P_r(d_r; \mu) = U_r(d_r) - \mu d_r. \tag{3.87}$$

Notice that in the previous expression, each user is acting as a price taker. We expect that *marginal cost pricing* would again yield an optimal solution to *SYSTEM*, i.e., that choosing $\mu(\mathbf{d}) = p(\sum_r d_r)$ would lead users to maximize aggregate surplus at a competitive equilibrium. This is formalized in the following proposition, proven using methods similar to the proof of Theorem 3.2.

Proposition 3.17
Suppose Assumptions 3.8 and 3.9 hold. There exists a competitive equilibrium*, that is, a*

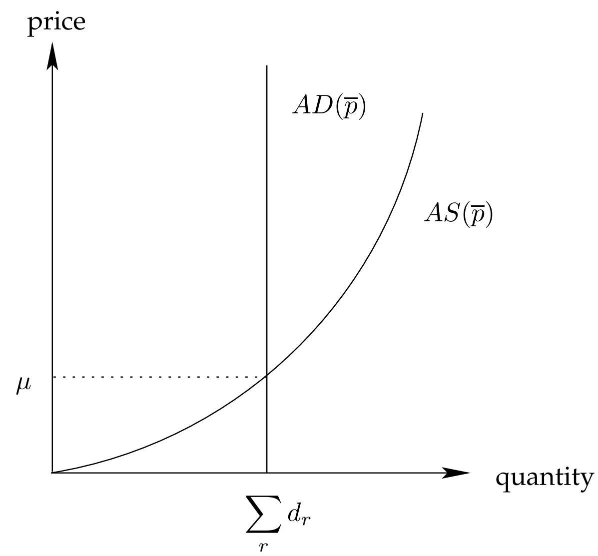

Figure 3-4. The market-clearing process for Cournot competition: Each consumer r chooses a quantity d_r, which maps to the inelastic demand function $D(p, d_r) = d_r$. This defines the aggregate demand function $AD(\overline{p}) = \sum_r D(\overline{p}, w_r) = \sum_r d_r$. The aggregate supply function is $AS(\overline{p}) = p^{-1}(\overline{p})$. The price μ is chosen so that supply equals demand, i.e., so that $\sum_r d_r = AD(\mu) = AS(\mu) = p^{-1}(\mu)$; in other words, $\mu = p(\sum_r d_r)$.

vector $\mathbf{d}$ *and a scalar* μ *such that* $\mu = p(\sum_r d_r)$*, and:*

$$P_r(d_r; \mu) = \max_{\overline{d}_r \geq 0} P_r(\overline{d}_r; \mu), \quad r = 1, \ldots, R. \tag{3.88}$$

Any such vector $\mathbf{d}$ *is an optimal solution to SYSTEM. If the functions* U_r *are strictly concave, such a vector* $\mathbf{d}$ *is unique as well.*

Observe that we can again interpret this process as a market-clearing process (see Section 3.1). Each user r chooses a parameter d_r; the demand function of user r as a function of price is then $D(p, d_r) = d_r$, so that demand is independent of price. The resource manager then chooses a price μ so that aggregate demand equals supply, where supply is given by the function $S(\mu) = p^{-1}(\mu)$; see Figure 3-4. Thus the price μ satisfies $\sum_r D(\mu, d_r) = S(\mu)$; inverting, this gives precisely the equation $\mu = p(\sum_r d_r)$.

Proposition 3.17 shows that with an appropriate choice of price function, and under the assumption that the users of the link behave as price takers, there exists a vector of rates $\mathbf{d}$ where all users have optimally chosen their d_r, with respect to the given price $\mu = p(\sum_r d_r)$; and at this "equilibrium," the aggregate surplus is maximized. However, when the price taking assumption is violated, the model changes into a game and the guarantee of Proposition 3.17 is no longer valid.

Consider, then, an alternative model where the users of a single link are price anticipating, rather than price takers, and play a Cournot game to acquire a share of the link. We use the notation $\mathbf{d}_{-r}$ to denote the vector of all rates chosen by users other than r; i.e., $\mathbf{d}_{-r} = (d_1, d_2, \ldots, d_{r-1}, d_{r+1}, \ldots, d_R)$. Then given $\mathbf{d}_{-r}$, each user r chooses $d_r \geq 0$ to maximize:

$$Q_r(d_r; \mathbf{d}_{-r}) = U_r(d_r) - d_r p \left(\sum_s d_s \right). \tag{3.89}$$

The payoff function Q_r is similar to the payoff function P_r, except that the user now anticipates that the network will set the price according to $p(\sum_s d_s)$. A *Nash equilibrium* of the game defined by $(Q_1, \ldots, Q_R)$ is a vector $\mathbf{d} \geq 0$ such that for all r:

$$Q_r(d_r; \mathbf{d}_{-r}) \geq Q_r(\overline{d}_r; \mathbf{d}_{-r}), \quad \text{for all } \overline{d}_r \geq 0. \tag{3.90}$$

It is straightforward to show that a Nash equilibrium exists for this game, as we prove in the following result; see also [92].

Proposition 3.18
Suppose that Assumptions 3.8 and 3.9 hold. Then there exists a Nash equilibrium $\mathbf{d}$ *for the game defined by* $(Q_1, \ldots, Q_R)$.

Proof. We begin by observing that we may restrict the strategy space of each user r to a compact set, without loss of generality. We simply observe that for all d_r larger than some sufficiently large D_r, we will have $U_r(d_r) < d_r p(d_r)$, so that for any vector $\mathbf{d}_{-r}$ of rates chosen by other users, user r would always be better off choosing $d_r = 0$ rather than $d_r > D_r$. Thus, we may restrict the strategy space of user r to the compact interval $S_r = [0, D_r]$ without loss of generality.

Next, note that since p satisfies Assumption 3.9, $d_r p(\sum_s d_s)$ is convex in $d_r \geq 0$ for any value of $\mathbf{d}_{-r}$. This ensures Q_r is concave in $d_r \geq 0$ for all $\mathbf{d}_{-r}$.

The game defined by $(Q_1, \ldots, Q_R)$ together with the strategy spaces $(S_1, \ldots, S_R)$ is now a *concave R-person game*: each payoff function Q_r is continuous in the composite strategy vector $\mathbf{d}$, and concave in d_r; and the strategy space of each user r is a compact, convex, nonempty subset of $\mathbb{R}$. Applying Rosen's existence theorem [104] (proven using Kakutani's fixed point theorem), we conclude that a Nash equilibrium $\mathbf{d}$ exists for this game. □

Because the payoff Q_r is concave in d_r for fixed $\mathbf{d}_{-r}$, a vector $\mathbf{d}$ is a Nash equilibrium if and only if the following first order conditions are satisfied for each r, where

$f = \sum_s d_s$:

$$U_r'(d_r) \le p(f) + d_r \frac{\partial^+ p(f)}{\partial f}; \tag{3.91}$$

$$U_r'(d_r) \ge p(f) + d_r \frac{\partial^- p(f)}{\partial f}, \quad \text{if } d_r > 0. \tag{3.92}$$

We will use these conditions to investigate the efficiency loss when users are price anticipating. Before continuing, however, we note in the following proposition that if the price function p is differentiable, then either all Nash equilibria are optimal solutions to *SYSTEM*, or there exists a unique Nash equilibrium.

Proposition 3.19
Suppose that Assumptions 3.8 and 3.9 hold, and the price function p is differentiable. Then at least *one of the following holds:*

1. *All Nash equilibria of the game defined by $(Q_1, \ldots, Q_R)$ are also optimal solutions to SYSTEM; or*

2. *There exists a unique Nash equilibrium of the game defined by $(Q_1, \ldots, Q_R)$.*

Proof. Suppose there exists a Nash equilibrium $\mathbf{d}$ which is not an optimal solution to *SYSTEM*. Let $f = \sum_r d_r$. From (3.91)-(3.92), we know that $U_r'(d_r) = p(f) + d_r p'(f)$ for all r with $d_r > 0$. Now if $p'(f) = 0$, then $U_r'(d_r) = p(f)$ for all r with $d_r > 0$, while $U_r'(d_r) \le p(f)$ for all r with $d_r = 0$ (from (3.91)). These are the optimality conditions for *SYSTEM*; thus if $p'(f) = 0$, then $\mathbf{d}$ is in fact an optimal solution to *SYSTEM*. Thus we cannot have $p'(f) = 0$, so $p'(f) > 0$.

Now let $\hat{\mathbf{d}}$ be another Nash equilibrium, and let $\hat{f} = \sum_r \hat{d}_r$. Assume that $\hat{\mathbf{d}} \ne \mathbf{d}$. We first suppose that $\hat{f} \le f$, and choose a user r such that $\hat{d}_r < d_r$. In this case $d_r > 0$, so we have $U_r'(d_r) = p(f) + d_r p'(f)$. But now $\hat{f} \le f$, $\hat{d}_r < d_r$, $p'(f) > 0$, convexity of p, and concavity of U_r imply that $U_r'(\hat{d}_r) \ge U_r'(d_r) = p(f) + d_r p'(f) > p(\hat{f}) + \hat{d}_r p'(\hat{f})$, which contradicts (3.91).

So we can only have $\hat{f} > f$; in particular, note that this implies $p'(\hat{f}) \ge p'(f) > 0$ by convexity of p. But then interchanging the roles of $\hat{\mathbf{d}}$ and $\mathbf{d}$, we can apply the same argument as the preceding paragraph to arrive at a contradiction. Thus we must have had $\hat{\mathbf{d}} = \mathbf{d}$, i.e., the Nash equilibrium is unique. □

We will now use the conditions (3.91)-(3.92) to analyze the efficiency loss when users are price anticipating under Cournot competition. We first show in the following example that, in general, the efficiency loss may be arbitrarily high.

Example 3.4

Consider a price function p defined as follows:

$$p(f) = \begin{cases} af, & 0 \le f \le 1; \\ a + b(f-1), & f \ge 1. \end{cases}$$

Note that this yields:

$$C(f) = \begin{cases} \frac{1}{2}af^2, & 0 \le f \le 1; \\ \\ \frac{1}{2}a + a(f-1) + \frac{1}{2}b(f-1)^2, & f \ge 1. \end{cases}$$

We assume that $0 < a < 1/2$, and $b > 1$. We consider a game with $R = 2$ users where $U_1(d_1) = d_1$, and:

$$U_2(d_2) = a\left(2 - \frac{1-a}{b}\right) d_2.$$

In this case, note that aggregate surplus is maximized when $p(f^S) = 1$, i.e., when $f^S = 1 + (1-a)/b$; and furthermore, this rate should be allocated entirely to user 1. Thus the maximal aggregate surplus is $U_1(f^S) - C(f^S)$, or:

$$1 + \frac{1-a}{b} - \frac{1}{2}a - \frac{a(1-a)}{b} - \frac{(1-a)^2}{2b}. \tag{3.93}$$

On the other hand, we claim that the vector $\mathbf{d}$ defined by:

$$d_1 = \frac{1-a}{b};$$
$$d_2 = 1 - \frac{1-a}{b},$$

is a Nash equilibrium. Observe that $f = d_1 + d_2 = 1$, so $p(f) = a$. Furthermore, at $f = 1$, we have $\partial^+ p(f)/\partial f = b$, $\partial^- p(f)/\partial f = a$. It then follows that (3.91) is satisfied with equality by user 1, and (3.92) is satisfied with equality by user 2. Since $a < b$, these conditions are sufficient to ensure that $\mathbf{d}$ is a Nash equilibrium. Note that the aggregate surplus at this Nash equilibrium is $U_1(d_1) + U_2(d_2) - C(f)$, or:

$$\frac{1-a}{b} + a\left(2 - \frac{1-a}{b}\right)\left(1 - \frac{1-a}{b}\right) - \frac{1}{2}a.$$

Comparing this expression with (3.93), it is clear that in the limit where $a \to 0$ and $b \to \infty$, the Nash equilibrium aggregate surplus approaches zero, and the maximal aggregate surplus approaches 1; thus the ratio of Nash equilibrium aggregate surplus to the maximal aggregate surplus approaches zero.

Despite the preceding negative result, in the next section we prove a sequence of three results characterizing efficiency loss in more limited environments. We then also demonstrate an extension of Cournot competition to a network context in Section 3.5.2. This network model has the appealing feature that users only need to select rates on paths through the network, rather than bidding at individual links as in the model of Section 3.4.

■ 3.5.1 Models with Bounded Efficiency Loss

In this section we consider restricted models of Cournot competition which allow positive results in bounding efficiency loss when users are price anticipating. The first two of these results make restrictive assumptions on the users; the third result restricts the price function to be affine. We start with the following theorem.

Theorem 3.20
Suppose that $R = 1$, and user 1 has utility function U such that Assumption 3.8 holds; in addition, suppose that Assumption 3.9 holds. Suppose also that $U(0) \geq 0$. If d^S is an optimal solution to SYSTEM, and d maximizes $U(\overline{d}) - \overline{d}p(\overline{d})$ over $\overline{d} \geq 0$, then:

$$U(d) - C(d) \geq \left(\frac{2}{3}\right)\left(U(d^S) - C(d^S)\right). \tag{3.94}$$

Furthermore, this bound is tight, i.e., there exists a choice of U and C such that (3.94) *holds with equality.*

Proof. First suppose that $U'(d) \leq p(d)$. Now if $U'(d) < p(d)$, then $d = 0$ (from (3.91)-(3.92)), and in this case d is an optimal solution to *SYSTEM*. On the other hand, if $U'(d) = p(d)$, then d is again an optimal solution to *SYSTEM*. In either of these cases, the bound (3.94) holds; so we can assume without loss of generality that $U'(d) > p(d)$.

Furthermore, note that $U(d) - C(d) \geq U(d) - dp(d) \geq 0$, by convexity of C; thus, if $U(d^S) - C(d^S) = 0$, then we must also have $U(d) - C(d) = 0$ since d^S is an optimal solution to *SYSTEM*. So if $U(d^S) - C(d^S) = 0$, we again conclude that (3.94) holds. Thus we can assume without loss of generality that $U(d^S) - C(d^S) > 0$, and $U'(d) > p(d)$. Now, by arguing as in Step 1 of Theorem 3.8, we can show that the worst case occurs when the utility function U is linear: $U(\overline{d}) = \alpha\overline{d}$ for some $\alpha > 0$ with $\alpha > p(d)$. We make this assumption for the remainder of the proof.

Since $\alpha > p(d)$, we conclude that $d > 0$ (using (3.91)). Furthermore, by applying the fact that p is nondecreasing we conclude that $d < d^S$ (since $\alpha = p(d^S)$).

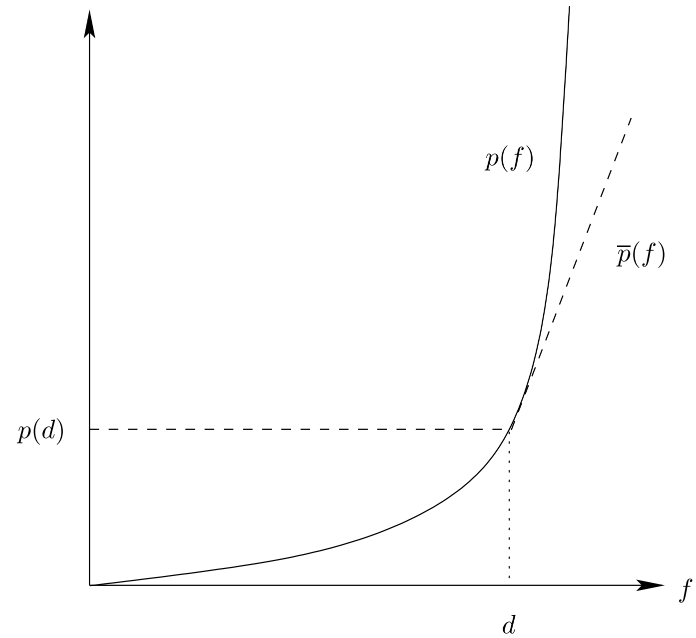

Figure 3-5. Proof of Theorem 3.20: Given a price function p (solid line) and Nash equilibrium rate d, a new price function $\overline{p}$ (dashed line) is defined according to (3.95).

We now argue as follows. Define a new price function $\overline{p}(f)$ according to:

$$\overline{p}(f) = \begin{cases} p(d), & f \leq d; \\ \\ p(d) + \left(\dfrac{\alpha - p(d)}{d}\right)(f-d), & f \geq d. \end{cases} \tag{3.95}$$

(See Figure 3-5 for an illustration.) Define $\overline{C}(f) = \int_0^f \overline{p}(z)\, dz$. Note that since $\alpha > p(d)$, $\overline{p}$ and $\overline{C}$ satisfy Assumption 3.9. We now claim that d maximizes $\alpha\overline{d} - \overline{d}\overline{p}(\overline{d})$ over $\overline{d} \geq 0$. We need only to check that (3.91)-(3.92) are satisfied. The condition (3.92) is satisfied since:

$$\alpha > p(d) = \overline{p}(d) + d\frac{\partial^- \overline{p}(d)}{\partial d}.$$

On the other hand, by definition of $\overline{p}(d)$ we have:

$$\frac{\partial^+ \overline{p}(d)}{\partial d} = \frac{\alpha - p(d)}{d},$$

so that (3.91) is satisfied with $\overline{p}$ in place of p. Thus d maximizes $\alpha\overline{d} - \overline{d}\overline{p}(\overline{d})$ over $\overline{d} \geq 0$. In particular, observe that $\alpha d - \overline{C}(d) = (\alpha - p(d))d$.

The optimal solution to *SYSTEM* when the utility function is U and the price function is $\overline{p}$ is given by solving $\alpha = \overline{p}(\overline{d}^S)$, which yields $\overline{d}^S = 2d$, and aggregate surplus $U(2d) - \overline{C}(2d) = 3(\alpha - p(d))d/2$. Thus the ratio of Nash equilibrium aggregate surplus to maximal aggregate surplus is $2/3$. To complete the proof of the theorem, therefore, it suffices to show that:

$$\frac{\alpha d - C(d)}{\alpha d^S - C(d^S)} \geq \frac{\alpha d - \overline{C}(d)}{\max_{\overline{d} \geq 0} \left(\alpha \overline{d} - \overline{C}(\overline{d})\right)}.$$

We argue as follows. Define an intermediate price function $\hat{p}(f)$ as follows:

$$\hat{p}(f) = \begin{cases} p(d), & f \leq d; \\ p(f), & f \geq d. \end{cases}$$

Define $\hat{C}(f) = \int_0^f \hat{p}(z)\ dz$; then it is straightforward to check that $\hat{p}$ satisfies Assumption 3.9. Since p is nondecreasing (Assumption 3.9), we have $p(d) \geq p(f)$ for $f \leq d$; and thus if we define $\Delta = \hat{C}(d) - C(d)$, then $\Delta \geq 0$. Furthermore, since we have already shown that $d^S > d$, we have $\hat{C}(d^S) = C(d^S) + \Delta$, or rearranging, $C(d^S) = \hat{C}(d^S) - \Delta$. Because $\alpha > p(d)$, we have $0 < \alpha d - dp(d) = \alpha d - \hat{C}(d) \leq \alpha d^S - \hat{C}(d^S)$ (where the latter inequality follows since d^S is an optimal solution to *SYSTEM* even when the cost function is $\hat{C}$). Thus we have:

$$\begin{aligned} \frac{\alpha d - C(d)}{\alpha d^S - C(d^S)} &= \frac{\alpha d - \hat{C}(d) + \Delta}{\alpha d^S - \hat{C}(d^S) + \Delta} \\ &\geq \frac{\alpha d - \hat{C}(d)}{\alpha d^S - \hat{C}(d^S)}. \end{aligned} \tag{3.96}$$

We now observe that $\hat{C}(d) = dp(d) = \overline{C}(d)$, so the numerator in the last expression is $\alpha d - \hat{C}(d) = \alpha d - \overline{C}(d)$. On the other hand, from (3.92) it follows that:

$$\alpha = U'(d) \leq p(d) + d\frac{\partial^+ p(d)}{\partial d}.$$

Rearranging, we conclude that:

$$\frac{\partial^+ \overline{p}(d)}{\partial d} = \frac{\alpha - p(d)}{d} \leq \frac{\partial^+ p(d)}{\partial d} = \frac{\partial^+ \hat{p}(d)}{\partial d}.$$

Thus since $\hat{p}$ is convex, we have $\hat{p}(f) \geq \overline{p}(f)$ for $f \geq d$; on the other hand, we have $\hat{p}(f) = \overline{p}(f)$ for $f \leq d$. Since $d^S > d$, we have $\hat{C}(d^S) \geq \overline{C}(d^S)$, so that $\alpha d^S - \hat{C}(d^S) \leq$

$\alpha d^S - \overline{C}(d^S) \leq \max_{\overline{d} \geq 0}(\alpha \overline{d} - \overline{C}(\overline{d}))$. Combining this inequality with (3.96) yields:

$$\frac{\alpha d - C(d)}{\alpha d^S - C(d^S)} \geq \frac{\alpha d - \overline{C}(d)}{\max_{\overline{d} \geq 0}\left(\alpha \overline{d} - \overline{C}(\overline{d})\right)} = \frac{2}{3},$$

as required. □

The preceding theorem considered a single user; in the next theorem, we consider a model where multiple users share the same utility function.

Theorem 3.21
Suppose that $R \geq 1$ users share the same utility function $U_r = U$, such that Assumption 3.8 holds; in addition, suppose that Assumption 3.9 holds, and that p is differentiable. Suppose also that $U(0) \geq 0$. If $\mathbf{d}^S$ is an optimal solution to SYSTEM, and $\mathbf{d}$ is a Nash equilibrium of the game defined by $(Q_1, \ldots, Q_R)$, then

$$\sum_r U(d_r) - C\left(\sum_r d_r\right) \geq \left(\frac{2R}{2R+1}\right)\left(\sum_r U(d_r^S) - C\left(\sum_r d_r^S\right)\right). \qquad (3.97)$$

In the special case where $U(d_r) = \alpha d_r$ for $\alpha > 0$, the preceding result holds even if p is not differentiable.

Proof. We start by assuming that p is differentiable. If all Nash equilibria are also optimal solutions to *SYSTEM*, then the theorem is trivially true; thus we assume without loss of generality that $\mathbf{d}$ is a Nash equilibrium which does not solve *SYSTEM*. In this case, from Proposition 3.19, we know that $\mathbf{d}$ is in fact the unique Nash equilibrium. Let $f = \sum_r d_r$; then by symmetry (since all users have the same utility function), the unique Nash equilibrium must be $d_r = f/R$ for all r. Furthermore, by symmetry, since $\sum_r U(d_r) - C(\sum_r d_r)$ is concave in d_r, there exists a symmetric optimal solution $\mathbf{d}^S$ to *SYSTEM*; i.e., letting $f^S = \sum_r d_r^S$, we have $d_r^S = f^S/R$.

We use the same basic argument as in the proof of Theorem 3.20. As in the opening of that proof, we can assume without loss of generality that $U'(d_r) > p(f)$ for all r, and $\sum_r U(d_r^S) - C(f^S) > 0$. In this case, by using an argument similar to the proof of Step 1 of Theorem 3.8, it follows that linear utility functions are the worst case. In fact, we conclude something stronger: the worst case occurs when all users have the *same* linear utility function, since $U'(d_r) = U'(d_s) = U'(f/R)$ for $r \neq s$. Thus, for the remainder of the proof, we assume without loss of generality that $U(\overline{d}) = \alpha \overline{d}$ for $\alpha > 0$, where $\alpha > p(f)$ at the Nash equilibrium. A key observation we make is that since $\alpha > p(f)$ at the Nash equilibrium, while $\alpha = p(f^S)$ at the optimal solution to *SYSTEM*, and p is nondecreasing, we must have $f^S > f$. In addition, since $U'(d_r) > p(f)$ for all r, we must have $d_r > 0$ for all r (from (3.91)), so that $f > 0$.

We now argue exactly as in the proof of Theorem 3.20. We define a new price function $\overline{p}$ as follows:

$$\overline{p}(\overline{f}) = \begin{cases} p(f), & \overline{f} \le f; \\ \\ p(f) + \left(\dfrac{(\alpha - p(f))R}{f}\right)(\overline{f} - f), & \overline{f} \ge f. \end{cases}$$

(Note the similarity with the definition (3.95) of $\overline{p}$ in the proof of Theorem 3.20.) We also define the intermediate price function $\hat{p}$ as follows:

$$\hat{p}(\overline{f}) = \begin{cases} p(f), & \overline{f} \le f; \\ p(\overline{f}), & \overline{f} \ge f. \end{cases}$$

Define the associated cost functions $\overline{C}(\overline{f}) = \int_0^{\overline{f}} \overline{p}(z)\ dz$, and $\hat{C}(\overline{f}) = \int_0^{\overline{f}} \hat{p}(z)\ dz$. We now use an argument similar to the proof of Theorem 3.20. If we define $\Delta = \hat{C}(f) - C(f)$, then $\Delta \ge 0$; and since $f^S > f$, we have $\hat{C}(f^S) = C(f^S) + \Delta$. Because $\alpha > p(f)$, we have $0 < \alpha f - fp(f) = \alpha f - \hat{C}(f) \le \alpha f^S - \hat{C}(f^S)$ (where the latter inequality follows since $\mathbf{d}^S$ is an optimal solution to *SYSTEM* even when the cost function is $\hat{C}$). Thus we have:

$$\begin{aligned} \frac{\alpha f - C(f)}{\alpha f^S - C(f^S)} &= \frac{\alpha f - \hat{C}(f) + \Delta}{\alpha f^S - \hat{C}(f^S) + \Delta} \\ &\ge \frac{\alpha f - \hat{C}(f)}{\alpha f^S - \hat{C}(f^S)}. \end{aligned} \tag{3.98}$$

We now observe that $\hat{C}(f) = fp(f) = \overline{C}(f)$, so the numerator in the last expression is $\alpha f - \hat{C}(f) = \alpha f - \overline{C}(f)$. On the other hand, since $\mathbf{d}$ is a Nash equilibrium, p is differentiable, and $d_r = f/R > 0$ for all r, from (3.91)-(3.92) we must have:

$$\alpha = p(f) + \frac{fp'(f)}{R}.$$

From this we conclude that:

$$\frac{\partial^+ \overline{p}(f)}{\partial f} = \frac{(\alpha - p(f))R}{f} = p'(f) = \frac{\partial^+ \hat{p}(f)}{\partial f}. \tag{3.99}$$

Since p is convex (Assumption 3.9), we know that $\hat{p}$ is convex; and thus we may conclude that $\hat{p}(\overline{f}) \ge \overline{p}(\overline{f})$ for $\overline{f} \ge f$. On the other hand, $\hat{p}(\overline{f}) = \overline{p}(\overline{f})$ for $\overline{f} \le f$. We conclude that $\hat{C}(\overline{f}) \ge \overline{C}(\overline{f})$ for $\overline{f} \ge f$. Since $f^S > f$, we have $\alpha f^S - \hat{C}(f^S) \le \alpha f^S - \overline{C}(f^S) \le$

$\max_{\overline{f}\geq 0}(\alpha\overline{f} - \overline{C}(\overline{f}))$. Thus:

$$\frac{\alpha f - C(f)}{\alpha f^S - C(f^S)} \geq \frac{\alpha f - \overline{C}(f)}{\max_{\overline{f}\geq 0}\left(\alpha\overline{f} - \overline{C}(\overline{f})\right)}. \tag{3.100}$$

We now explicitly compute the right hand side. The numerator is $(\alpha - p(f))f$. The maximum in the denominator is achieved when $\overline{p}(\overline{f}) = \alpha$, i.e., when $\overline{f} = f + f/R$. It is straightforward to check that in this case the denominator is equal to $(\alpha - p(f)(f + f/2R)$. Thus, the ratio on the right hand side of (3.100) is exactly equal to $2R/(2R+1)$.

Note that the only step in the above argument that required differentiability of the price function is the reduction to linear utility functions where each user has exactly the same slope. If we assume to begin with that each user has utility function $U(d) = \alpha d$, then we do not require differentiability of p. To see this, let p be any price function satisfying Assumption 3.9, and let $\mathbf{d}$ be any Nash equilibrium (note that it is no longer necessarily unique); let $f = \sum_r d_r$. As in the preceding development, we can assume without loss of generality that $\alpha > p(f)$, $\alpha f^S - C(f^S) > 0$, and $f^S > f > 0$. We now recall the optimality condition (3.91) for each user r:

$$\alpha \leq p(f) + d_r\frac{\partial^+ p(f)}{\partial f}.$$

If we consider a user r such that $d_r \leq f/R$ (at least one such user exists), then the preceding inequality implies:

$$\frac{(\alpha - p(f))R}{f} \leq \frac{\partial^+ p(f)}{\partial f}.$$

It is now possible to verify that the proof of the theorem holds as before if we replace (3.99) by the following inequality:

$$\frac{\partial^+\overline{p}(f)}{\partial f} = \frac{(\alpha - p(f))R}{f} \leq \frac{\partial^+ p(f)}{\partial f} = \frac{\partial^+\hat{p}(f)}{\partial f}.$$

Thus the theorem holds in this case, even if the price function p is not necessarily differentiable. □

Note that although a tightness result is not claimed in the theorem, such a result may be established by considering a limit of differentiable price functions which approach the worst case price function $\overline{p}$ defined in the theorem. However, defining such price functions requires additional technical complexity, and does not yield additional insight; thus the argument is omitted.

Note that the preceding theorem also implicitly yields a competitive limit theorem [82], since as $R \to \infty$ the efficiency loss approaches zero. Indeed, this result is to be expected, since the users are assumed to be symmetric; thus in the limit of many users no single user should have a significant impact on the market-clearing price.

Theorems 3.20 and 3.21 present bounds on efficiency loss under various restrictions on utility functions and the price function p. We note that as was done in the shift from Assumption 3.1 to Assumption 3.5, these results would continue to hold even if the utility functions were not necessarily differentiable (as we required in Assumption 3.8). Differentiability of the utility function only eases the presentation of the technical arguments, but is not essential to the results.

By contrast, differentiability of the price function p is essential to the proof of Theorem 3.21. In particular, in considering the statements of Theorems 3.20 and 3.21, one might expect a more general result to hold: if R users share the same utility function U and Assumption 3.8 is satisfied, and the price function p satisfies Assumption 3.9 (but is not necessarily differentiable), then the efficiency loss is no more than $1/(2R+1)$ when users are price anticipating. Such a result would be a generalization of both Theorem 3.20 and Theorem 3.21.

However, the efficiency loss can be arbitrarily high if the price function is not differentiable, even if all users share the same utility function. The main reason for this negative result is that when the price function is not differentiable, there exist Nash equilibria which are not symmetric among the players; this symmetry plays a key role in the proof of Theorem 3.21. We present an example here of such a situation.

Example 3.5

Let the number of users be $R > 1$, and let the price function be $p(f) = (f-1)^+$. Let $C(f) = \int_0^f p(z)\, dz$ be the associated cost function; note that p and C satisfy Assumption 3.9. Define $\alpha = 1/R^2$ and $\hat{d} = (\alpha+1)/R$. We then define the piecewise linear utility function U as follows:

$$U(d) = \begin{cases} \alpha d, & \text{if } d \le \hat{d}; \\ \alpha \hat{d}, & \text{if } d \ge \hat{d}. \end{cases}$$

Then U is concave and continuous. Note that U is not differentiable, but as discussed above, this feature is inessential to the argument; a similar example can be constructed with a differentiable utility function U, at considerably higher technical expense.

We now claim that if $d_r^S = \hat{d}$ for all r, then $\mathbf{d}^S$ is a optimal solution to *SYSTEM*. To see this, note that $f^S = \sum_r d_r^S = R\hat{d} = \alpha + 1$; and thus $p(f^S) = \alpha$. On the other hand, we have:

$$\frac{\partial^+ U(d_r^S)}{\partial d_r} = \frac{\partial^+ U(\hat{d})}{\partial d} = 0 < \alpha = p(f^S);$$
$$\frac{\partial^- U(d_r^S)}{\partial d_r} = \frac{\partial^- U(\hat{d})}{\partial d} = \alpha = p(f^S).$$

These are necessary and sufficient optimality conditions for $\mathbf{d}^S$ to be a optimal solution to *SYSTEM*, as required. Note that the aggregate surplus at this solution is $\sum_r U(d_r^S) - C(f^S) = R\alpha\hat{d} - \alpha^2/2 = \alpha^2/2 + \alpha$.

Next, let $d_r = \alpha$ for $r = 2, \ldots, R$, and $d_1 = 1 - (R-1)\alpha$. Note that $f = \sum_r d_r = 1$, and thus $p(f) = 0$ and:

$$\frac{\partial^- p(f)}{\partial f} = 0; \quad \frac{\partial^+ p(f)}{\partial f} = 1.$$

We claim $\mathbf{d}$ is a Nash equilibrium; note that $\mathbf{d}$ is not symmetric among the players. Using the definitions of $\hat{d}$ and α, it is straightforward to establish that $0 < \alpha < \hat{d} < 1 - (R-1)\alpha$ as long as $R > 1$. Thus, in particular, U is differentiable at d_r for all r, and $U'(d_1) = 0$, while $U'(d_r) = \alpha$. Now we observe that:

$$p(f) + d_1 \frac{\partial^- p(f)}{\partial f} = 0 = U'(d_1) < 1 - (R-1)\alpha = p(f) + d_1 \frac{\partial^+ p(f)}{\partial f};$$

$$p(f) + d_r \frac{\partial^- p(f)}{\partial f} = 0 < U'(d_r) = \alpha = p(f) + d_r \frac{\partial^+ p(f)}{\partial f}, \quad r = 2, \ldots, R.$$

These conditions are identical to the necessary and sufficient conditions (3.91)-(3.92), so we conclude $\mathbf{d}$ is a Nash equilibrium. At this Nash equilibrium, the aggregate surplus is $\sum_r U(d_r) - C(f) = \alpha\hat{d} + (R-1)\alpha^2$. If we now substitute $\alpha = 1/R^2$ and $\hat{d} = (\alpha + 1)/R = 1/R^3 + 1/R$, the ratio of Nash equilibrium aggregate surplus to the maximal aggregate surplus reduces to:

$$\frac{1/R^5 + 1/R^3 + (R-1)/R^4}{1/(2R^4) + 1/R^2}.$$

As $R \to \infty$, the preceding ratio approaches zero. □

The preceding example highlights an important issue in market modeling: results on the performance of the market can be very sensitive under assumptions of symmetry among the participants. In particular, one might expect that little difference exists in market performance whether the price function is differentiable or not; nevertheless, the preceding example shows that efficiency loss can become arbitrarily high if the price function is not differentiable. To avoid such singular effects, we now search instead for a result that holds regardless of the utility functions of the users. Of course, such a result cannot hold for all price functions. In particular, we prove in the following theorem that if the price function is linear, the resulting efficiency loss is no more than $1/3$ of the maximal aggregate surplus, regardless of the utility functions of the users.

Theorem 3.22
Suppose that Assumption 3.8 holds, and that $p(f) = af$ for some $a > 0$. Suppose also that $U_r(0) \geq 0$ for all r. If $\mathbf{d}^S$ is any optimal solution to SYSTEM, and $\mathbf{d}$ is any Nash equilibrium of the game defined by $(Q_1, \ldots, Q_R)$, then:

$$\sum_r U_r(d_r) - C\left(\sum_r d_r\right) \geq \left(\frac{2}{3}\right)\left(\sum_r U_r(d_r^S) - C\left(\sum_r d_r^S\right)\right). \tag{3.101}$$

Furthermore, this bound is tight: for every $\delta > 0$, there exists a choice of R and a choice of (linear) utility functions U_r, $r = 1, \ldots, R$ such that a Nash equilibrium $\mathbf{d}$ exists with:

$$\sum_r U_r(d_r) - C\left(\sum_r d_r\right) \leq \left(\frac{2}{3} + \delta\right)\left(\sum_r U_r(d_r^S) - C\left(\sum_r d_r^S\right)\right). \tag{3.102}$$

Proof. We first note that Steps 1 and 2 of the proof of Theorem 3.8 hold in this setting as well; i.e., we may assume without loss of generality that $U_r(d_r) = \alpha_r d_r$, where $0 < \alpha \leq 1 = \max_r \alpha_r$, and that the aggregate Nash equilibrium rate is equal to 1. Note that after rescaling, as required in those steps, the new price function p is still linear but may have a different slope. Since the price function is fixed as $p(f) = af$, the maximal aggregate surplus is achieved when $p(f^S) = 1$, i.e., when $f^S = 1/a$; and the maximal aggregate surplus is $1/a - (a/2)(1/a)^2 = 1/(2a)$.

Without loss of generality, we have restricted attention to situations where the aggregate Nash equilibrium rate is $f = 1$. We define $p = p(1) = a$, $C = C(1) = a/2$, and $p' = p'(1) = a$. Since the maximal aggregate surplus is fixed as $1/(2a)$, by (3.91)-(3.92) the worst case game is identified by solving the following optimization problem (with unknowns $d_1, \ldots, d_R, \alpha_2, \ldots, \alpha_R$):

$$\text{minimize} \quad d_1 + \sum_{r=2}^{R} \alpha_r d_r - C \tag{3.103}$$

$$\text{subject to} \quad \alpha_r \leq p + d_r p', \quad r = 1, \ldots, R; \tag{3.104}$$

$$\alpha_r \geq p + d_r p', \quad \text{if } d_r > 0, \;\; r = 1, \ldots, R; \tag{3.105}$$

$$\sum_{r=1}^{R} d_r = 1; \tag{3.106}$$

$$0 < \alpha_r \leq 1, \quad r = 2, \ldots, R; \tag{3.107}$$

$$d_r \geq 0, \quad r = 1, \ldots, R. \tag{3.108}$$

The objective function is the aggregate surplus given a Nash equilibrium allocation **d**. The conditions (3.104)-(3.105) are equivalent to the Nash equilibrium conditions established in (3.91)-(3.92). The constraint (3.106) ensures that the total allocation made at the Nash equilibrium is equal to 1, and the constraint (3.107) follows from Step 1 of the proof of Theorem 3.8. The constraint (3.108) ensures the rate allocated to each user is nonnegative.

Our approach is to solve this problem through a sequence of reductions. As in the proof of Theorem 3.8, it follows that we may assume without loss of generality that the constraint (3.105) holds with equality for all users $r = 2, \ldots, R$. Furthermore, it follows from (3.107) together with (3.104)-(3.105) that a candidate solution satisfying (3.106) can only exist if $d_1 > 0$, in which case we have $1 = p + d_1 p'$, so that $d_1 = (1-p)/p'$. In particular, we conclude immediately that for a feasible solution to exist, we must have $0 < (1-p)/p' \leq 1$. This yields the following reduced optimization problem:

$$\text{minimize} \quad \frac{1-p}{p'} + \sum_{r=2}^{R} (p + d_r p') d_r - C \tag{3.109}$$

$$\text{subject to} \quad \sum_{r=2}^{R} d_r = 1 - \frac{1-p}{p'}; \tag{3.110}$$

$$d_r \leq \frac{1-p}{p'}, \quad r = 2, \ldots, R; \tag{3.111}$$

$$d_r \geq 0, \quad r = 2, \ldots, R. \tag{3.112}$$

The objective function (3.109) is equivalent to (3.103) upon substitution for α_r (assuming equality in (3.105)) and d_1 (also by requiring equality in (3.105)). The constraint (3.110) is equivalent to the allocation constraint (3.106); and the constraint (3.111) ensures $\alpha_r \leq 1$, as required in (3.107).

The resulting problem is symmetric in the users $r = 2, \ldots, R$. It is clear that a feasible solution exists if and only if:

$$\frac{1}{R} \leq \frac{1-p}{p'} \leq 1. \tag{3.113}$$

In this case the following symmetric solution is feasible:

$$d_r = \frac{1-(1-p)/p'}{R-1}.$$

(Note that $d_r \geq 0$ because we know that $(1-p)/p' \leq 1$ for a feasible solution to exist.) By an argument similar to the proof of Step 3 of Theorem 3.8, we see that the worst case occurs as $R \to \infty$, and in this case the optimal objective value (3.109) becomes:

$$\frac{1-p}{p'} + p\left(1 - \frac{1-p}{p'}\right) - C = \frac{1}{a} - 1 + a\left(2 - \frac{1}{a}\right) - \frac{a}{2}.$$

Furthermore, the feasibility requirements (3.113) on p and p' become $0 < (1-p)/p' \leq 1$; upon substituting for p and p', these become $1/2 \leq a \leq 1$.

Recall that the maximal aggregate surplus is $1/(2a)$. Thus, the worst case ratio is identified by the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & \frac{1/a - 1 + a\,(2 - 1/a) - a/2}{1/2a} \\ \text{subject to} \quad & 1/2 \leq a \leq 1. \end{aligned}$$

It is straightforward to establish that the minimum value of this optimization problem occurs at $a = 2/3$, and the minimum objective value is equal to $2/3$. This establishes (3.101).

We now show (3.102), for a fixed price function $p(f) = af$ with $a > 0$. To see this, choose the utility functions so that $U_1(d_1) = 3ad_1/2$, and $U_r(d_r) = a(1+\overline{d})d_r$, where $\overline{d} = 1/(2(R-1))$. It is then straightforward to check that for sufficiently large R, if $d_1 = 1/2$ and $d_r = \overline{d}$ for $r = 2, \ldots, R$, the allocation $\mathbf{d}$ is a Nash equilibrium. Furthermore, the maximum aggregate surplus is achieved by choosing f^S so that $3a/2 = p(f^S) = af^S$, so $f^S = 3/2$. Thus the ratio of Nash equilibrium aggregate surplus to maximal aggregate surplus is:

$$\frac{(3a/2)(1/2) + a(1+\overline{d})/2 - a/2}{(3a/2)(3/2) - (a/2)(3/2)^2} = \frac{3/2 + \overline{d}}{9/4}.$$

Now as $R \to \infty$, this ratio approaches $2/3$, as required. □

Note that while the proof makes it appear as if the worst case occurs when the price function has slope $2/3$, in fact by an appropriate choice of utility functions the worst case efficiency loss is always *exactly* $1/3$ for *any* linear price function.

■ 3.5.2 General Networks

It is straightforward to extend Cournot competition to a network game of the form described in Section 3.4, where the strategy of a user r is a vector $\mathbf{x}_r = (x_{jr}, j \in J)$ of rates desired from the resources $j \in J$, and the payoff to user r is:

$$T_r(\mathbf{x}_r; \mathbf{x}_{-r}) = U_r(d_r(\mathbf{x}_r)) - \sum_j x_{jr} p_j \left(\sum_s x_{js} \right).$$

With this formulation, it can be shown that Nash equilibria exist, and the result of Theorem 3.22 continues to hold; the proof techniques are identical to those in Section 3.4. (Note that Theorem 3.21 may not extend, since users do not in general share the same topology, and thus have different *effective* utility functions $U_r(d_r(\cdot))$.)

The extension of the Cournot game to general networks has an additional attractive feature: it is equivalent to a game where users choose only the rate they desire on available *paths*, rather than choosing rates on a per-link basis. To formalize this notion, consider an alternative game where the strategy of a user r is $\mathbf{y}_r = (y_q, q \in r) \geq 0$, the rates user r wishes to send on each of the paths available to him. Define the payoff to user r as follows:

$$\overline{T}_r(\mathbf{y}_r; \mathbf{y}_{-r}) = U_r \left(\sum_{q \in r} y_q \right) - \sum_{q \in r} y_q \sum_{j \in q} p_j \left(\sum_{\overline{q}: j \in \overline{q}} y_{\overline{q}} \right).$$

The first term is the utility to user r; the last term is the total payment user r makes to the network. The expression $\sum_{j \in q} p_j (\sum_{\overline{q}: j \in \overline{q}} y_{\overline{q}})$ is the total price to user r of the path p. Note that this is a much more natural game for the users to play: it is reasonable to expect users to choose rates on a per path basis, based on observation of the prices of each of those paths.

We have the following theorem.

Theorem 3.23
Suppose that Assumption 3.5 holds, and that each price function p_j and cost function C_j satisfy Assumption 3.9. Let $\mathbf{y}$ be a Nash equilibrium of the game defined by $(\overline{T}_1, \ldots, \overline{T}_R)$, and define $x_{jr} = \sum_{q \in r: j \in q} y_q$. Then $\mathbf{x}$ is a Nash equilibrium of the game defined by $(T_1, \ldots, T_R)$.

Conversely, let $\mathbf{x}$ be a Nash equilibrium of the game defined by $(T_1, \ldots, T_R)$, and let $\mathbf{y}_r$ be any optimal solution to (3.68)-(3.70) *with $\overline{\mathbf{x}}_r = \mathbf{x}_r$. Then $\mathbf{y}$ is a Nash equilibrium of the game defined by $(\overline{T}_1, \ldots, \overline{T}_R)$.*

Proof. Suppose first that the vector $\mathbf{y}$ is a Nash equilibrium of the game defined by $(\overline{T}_1, \ldots, \overline{T}_R)$, and define $x_{jr} = \sum_{q \in r : j \in q} y_q$. By definition it follows that:

$$\sum_{q \in r} y_q \sum_{j \in q} p_j \left(\sum_{\overline{q} : j \in \overline{q}} y_{\overline{q}} \right) = \sum_j x_{jr} p_j \left(\sum_s x_{js} \right). \tag{3.114}$$

We now claim that $U_r(\sum_{q \in r} y_q) = U_r(d_r(\mathbf{x}_r))$. Clearly $\mathbf{y}_r$ is feasible for (3.68)-(3.70) with $\overline{\mathbf{x}}_r = \mathbf{x}_r$, so $\sum_{q \in r} y_q \le d_r(\mathbf{x}_r)$; and thus $U_r(\sum_{q \in r} y_q) \le U_r(d_r(\mathbf{x}_r))$. Suppose then that $U_r(\sum_{q \in r} y_q) < U_r(d_r(\mathbf{x}_r))$; this is only possible if $\sum_{q \in r} y_q < d_r(\mathbf{x}_r)$. But in this case if $\overline{\mathbf{y}}_r$ is an optimal solution to (3.68)-(3.70) with rate allocation $\mathbf{x}_r$, then $\overline{\mathbf{y}}_r$ is a profitable deviation for user r in the game defined by $(\overline{T}_1, \ldots, \overline{T}_R)$: the utility to user r strictly increases, while the payment made by user r does not increase. This is not possible, since $\mathbf{y}$ is a Nash equilibrium of the game defined by $(\overline{T}_1, \ldots, \overline{T}_R)$. We conclude that $U_r(\sum_{q \in r} y_q) = U_r(d_r(\mathbf{x}_r))$, and thus we have $T_r(\mathbf{x}_r; \mathbf{x}_{-r}) = \overline{T}_r(\mathbf{y}_r; \mathbf{y}_{-r})$ for all r.

Now suppose that $\mathbf{x}$ is not a Nash equilibrium of the game defined by $(T_1, \ldots, T_R)$; then there exists a user r such that $\overline{\mathbf{x}}_r \ne \mathbf{x}_r$ is a profitable deviation for user r. It is then straightforward to show that if $\overline{\mathbf{y}}_r$ is an optimal solution to (3.68)-(3.70), then $\overline{\mathbf{y}}_r$ is a profitable deviation for user r from $\mathbf{y}_r$, i.e., $\overline{T}_r(\overline{\mathbf{y}}_r; \mathbf{y}_{-r}) > \overline{T}_r(\mathbf{y}_r; \mathbf{y}_{-r})$. Thus $\mathbf{y}$ could not have been a Nash equilibrium for the game defined by $(\overline{T}_1, \ldots, \overline{T}_R)$, a contradiction. So we conclude $\mathbf{x}$ is a Nash equilibrium for the game defined by $(T_1, \ldots, T_R)$, as required.

Conversely, suppose $\mathbf{x}$ is a Nash equilibrium for the game defined by $(T_1, \ldots, T_R)$, and let $\mathbf{y}_r$ be any optimal solution to (3.68)-(3.70) for each r; then $\sum_{q \in r} y_q = d_r(\mathbf{x}_r)$. From the constraint (3.69), we have for each r that $\sum_{q \in r : j \in q} y_q \le x_{jr}$, from which it follows that for each r:

$$\sum_{q \in r} y_q \sum_{j \in q} p_j \left(\sum_{\overline{q} : j \in \overline{q}} y_{\overline{q}} \right) \le \sum_j x_{jr} p_j \left(\sum_s x_{js} \right).$$

Thus $\overline{T}_r(\mathbf{y}_r; \mathbf{y}_{-r}) \ge T_r(\mathbf{x}_r; \mathbf{x}_{-r})$. So now suppose that $\mathbf{y}$ is not a Nash equilibrium of the game defined by $(\overline{T}_1, \ldots, \overline{T}_R)$; then there exists a user r and a vector $\overline{\mathbf{y}}_r$ which is a profitable deviation for user r from $\mathbf{y}_r$. But then let $\overline{x}_{jr} = \sum_{q \in r : j \in q} \overline{y}_q$. It then follows that $\overline{\mathbf{x}}_r$ is a profitable deviation for user r from $\mathbf{x}_r$, so that $\mathbf{x}$ could not have been a Nash equilibrium for the game defined by $(T_1, \ldots, T_R)$—a contradiction. Thus $\mathbf{y}$ is a Nash equilibrium for the game defined by $(\overline{T}_1, \ldots, \overline{T}_R)$, as required. □

The preceding theorem shows that when users play a Cournot game in a general network, there is essentially no difference in the structure of the game whether they choose their rate demands at each link in the network, or along each path available through the network. Indeed, it is this equivalence which drives the proof of the theorem. From an architectural standpoint, such a theorem is very appealing, since the game where users select rates only along paths yields a much more natural and scalable network market mechanism. Formally, Theorem 3.23 allows us to conclude that the game defined by $(\overline{T}_1, \ldots, \overline{T}_R)$ inherits the mathematical properties of the game defined by $(T_1, \ldots, T_R)$; in particular, Nash equilibria exist for the game defined by $(\overline{T}_1, \ldots, \overline{T}_R)$, and the efficiency loss is no more than $1/3$ when users are price anticipating in the special case where p_j is linear for each j.

■ 3.6 Cournot Competition with Latency

Throughout this chapter, as well as Chapter 2, we have focused on users whose utility is determined only as a function of the data rate allocation. However, users of large networks (both communication networks and others) are not necessarily sensitive only to the amount of resources they are allocated; they may also care about other measures of performance. In this section, we consider a model where a user's utility is determined by both the rate allocation to that user, as well as the *latency* experienced by that user—i.e., the delay that the user's traffic suffers in traversing the network. For simplicity, we continue to focus on Cournot competition, so that the strategy space of the users is simply the data rate allocation they wish to receive. The network chooses a price per unit rate, and users choose their rate allocation to maximize their utility less their payment.

In the special case where the network does not charge any price, such a model bears close relation to the celebrated "selfish routing" model of Wardrop [10, 55, 61, 144]. That model initially considered a setting where users choose routes through a network to minimize their own latency; the rate users wish to send is assumed to be fixed, or inelastic. Our model is thus an extension to a setting where users' demands are elastic, rather than fixed. Roughgarden and Tardos [108] analyzed efficiency loss in the selfish routing model when users have fixed demands; in this section we will show that in general, efficiency loss can be arbitrarily high when users' demands are elastic (see Example 3.7).

We will investigate the model for a single link. Users have utility functions measured in monetary units, which depend on both allocated rate and experienced latency. We allow the possibility of a cost function at the link (which may represent a provisioning cost, for example), which is also measured in monetary units. Under simple conditions, we show there exists a aggregate surplus maximizing allocation; we also show that a competitive equilibrium exists, when users are both latency taking and price taking.[1] However, we then show that for a number of cases of interest, the efficiency loss between the aggregate surplus at the competitive equilibrium and the

[1] The term *latency taking* here is used in the same sense as price taking: users do not anticipate the effect of their actions on the latency they experience.

maximal aggregate surplus can be arbitrarily high. Note that in this setting, there is an efficiency loss even when users are latency taking and price taking; as we will identify, this efficiency loss occurs because the correct price is not being set for the users, even at the competitive equilibrium.

The formal specification of our model is as follows. We assume that each user r is endowed with a utility function $U_r(d_r, l_r)$, where d_r is the rate allocated to user r, and l_r is the latency experienced by user r. We will assume the link has a latency function $l(f)$, which specifies the delay experienced by a single unit of traffic on the link when the total rate through the link is f. We make the following assumptions on U_r and l.

Assumption 3.10
For each r, U_r is a concave and continuous function of the pair $(d_r, l_r) \geq 0$, such that U_r is nonincreasing in l_r, with $U_r(0, l_r) \geq 0$ for all l_r, and for any positive constant a, $U_r(d_r, ad_r) \to -\infty$ as $d_r \to \infty$. Furthermore, we assume the directional derivatives of U_r are continuous over the region $(d_r, l_r) \geq 0$; in particular, U_r is continuously differentiable for $d_r > 0$, $l_r > 0$.

Assumption 3.11
Over the domain $f \geq 0$, the function $l(f)$ is convex, strictly increasing, and continuous, with $l(0) = 0$ and $l(f) \to \infty$ as $f \to \infty$; and over the domain $f > 0$, the function $l(f)$ is continuously differentiable.

As an example, observe that Assumption 3.10 is satisfied by any separable utility function $U_r(d_r, l_r) = u_r(d_r) - d_r l_r$, where u_r is concave, nondecreasing, nonnegative, and differentiable. (Indeed, this is precisely the example considered in Example 3.7.) In addition, we will allow the possibility of a cost function $C(f)$, which may represent the monetary cost of provisioning the link. We will make the following assumption on C, which also allows the possibility that C is identically zero.

Assumption 3.12
There exists a continuous, nonnegative, and nondecreasing function $p(f)$ over $f \geq 0$ such that for $f \geq 0$:

$$C(f) = \int_0^f p(z)dz.$$

Thus $C(f)$ is convex and nondecreasing.

As in the previous sections, we will be interested in maximizing aggregate surplus. Formally, we define the *SYSTEM* problem for this setting as follows:

SYSTEM:

$$\text{maximize} \quad \sum_r U_r\left(d_r, l\left(\sum_s d_s\right)\right) - C\left(\sum_r d_r\right) \tag{3.115}$$

$$\text{subject to} \quad d_r \geq 0, \quad r = 1, \dots, R. \tag{3.116}$$

Observe that in this problem, each user r experiences a latency per unit rate given by $l_r = l(\sum_s d_s)$. Since the objective function is not separable over the users, it is not immediately obvious that the problem is convex; however, this is indeed the case, as demonstrated in the next proposition.

Proposition 3.24
Suppose that Assumptions 3.10-3.12 hold. Then the objective function (3.115) *is concave in the vector* $\mathbf{d}$*; and furthermore, there exists at least one optimal solution to SYSTEM.*

Proof. Fix two vectors $\mathbf{d}^1$, $\mathbf{d}^2$, and $\delta \in (0,1)$. We have:

$$\begin{aligned}
U_r&\left(\delta d_r^1 + (1-\delta)d_r^2, l\left(\delta \sum_s d_s^1 + (1-\delta)\sum_s d_s^2\right)\right) \\
&\geq U_r\left(\delta d_r^1 + (1-\delta)d_r^2, \delta l\left(\sum_s d_s^1\right) + (1-\delta) l\left(\sum_s d_s^2\right)\right) \\
&\geq \delta U_r\left(d_r^1, l\left(\sum_s d_s^1\right)\right) + (1-\delta)U_r\left(d_r^2, l\left(\sum_s d_s^2\right)\right),
\end{aligned}$$

where the first inequality follows because l is convex and U_r is nonincreasing in l_r; and the second inequality follows because U_r is concave. This establishes that U_r is concave as a function of $\mathbf{d}$, so that (3.115) is concave in $\mathbf{d}$.

Finally, to see that there exists at least one optimal solution, it suffices to show that $U_r(d_r, l(d_r)) \to -\infty$ as $d_r \to \infty$. To see this, fix $f_0 > 0$, and choose $a > 0$ such that $l(f) \geq af$ if $f \geq f_0$ (such a choice is possible given our assumptions on l). Then we have $U_r(d_r, l(d_r)) \leq U_r(d_r, ad_r)$ for $d_r \geq f_0$; and since $U_r(d_r, ad_r) \to -\infty$ as $d_r \to \infty$, we have $U_r(d_r, l(d_r)) \to -\infty$ as $d_r \to \infty$, as required. □

From the previous proposition, a vector $\mathbf{d}$ is an optimal solution to *SYSTEM* if and only if for all r we have:

$$\begin{aligned}
&\frac{\partial^+ U_r(d_r, l)}{\partial d_r} + l' \sum_s \frac{\partial^+ U_s(d_s, l)}{\partial l} \leq p; \\
&\frac{\partial^- U_r(d_r, l)}{\partial d_r} + l' \sum_s \frac{\partial^- U_s(d_s, l)}{\partial l} \geq p, , \quad \text{if } d_r > 0,
\end{aligned}$$

where $l = l(\sum_s d_s)$; $l' = l'(\sum_s d_s)$ (where we interpret $l'(0)$ as the right directional derivative of $l(0)$); and $p = p(\sum_s d_s)$. Intuitively, if all users act as latency takers and price takers and U_s is differentiable, then these optimality conditions suggest that the link manager should set a price given by $q = p - l' \sum_s \partial U_s(d_s, l)/\partial l$ (observe this price is positive, since the second term is negative). However, such a price would depend on the individual rate allocations made to the users, rather than on the aggregate rate allocation $\sum_s d_s$. For this reason, we restrict attention to pricing schemes where the resource manager chooses a price $t(\sum_s d_s)$ depending only on the aggregate rate allocation.

We now turn our attention to this model. Suppose that given a rate allocation $\mathbf{d}$, the resource manager charges a price $t(\sum_s d_s)$; we assume that the function $t(\cdot)$ is continuous and nonnegative. We then say that a vector $\mathbf{d}$ is a *competitive equilibrium* if for $l = l(\sum_s d_s)$, $t = t(\sum_s d_s)$, there holds:

$$d_r \in \arg\max_{\overline{d}_r \geq 0} \left[U_r(\overline{d}_r, l) - t\overline{d}_r \right]. \tag{3.117}$$

Observe that $U_r(d_r, l) - td_r$ is the payoff to user r when the latency is l, the price per unit rate is t, and the allocation to user r is d_r. Thus, user r chooses d_r to maximize this payoff, given a fixed latency l and price t; i.e., user r acts as a latency taker and a price taker. The next proposition shows there always exists at least one competitive equilibrium.

Proposition 3.25

Suppose that Assumptions 3.10-3.12 hold. Suppose also that $t(f)$ is continuous and nonnegative. Then there exists a competitive equilibrium $\mathbf{d}$.

Proof. In this case we cannot frame the competitive equilibrium as the optimal solution to an optimization problem. Instead, we use a fixed point approach. We start by showing that in searching for a competitive equilibrium, we may restrict the strategy space of each user to a compact set. To see this, note that since $U_r(d_r, ad_r) \to -\infty$ as $d_r \to \infty$ for any $a > 0$, we may argue as in the proof of Proposition 3.24 to show that $U_r(d_r, l(d_r)) \to -\infty$ as $d_r \to \infty$. In particular, for sufficiently large D_r, we will have $U_r(D_r, l(D_r)) < 0$. Thus if there exists a competitive equilibrium $\mathbf{d}$ with $d_r > D_r$, we will have $U_r(d_r, l) - td_r < 0$ (where $t = t(\sum_s d_s)$ and $l = l(\sum_s d_s)$), since l is increasing, and U_r is nonincreasing in l. But in this case $d_r = 0$ is a profitable deviation for user r, so $\mathbf{d}$ could not have been a competitive equilibrium. We thus restrict the strategy space of each user r to $[0, D_r]$.

Define for each user r:

$$BR_r(t, l) = \arg\max_{\overline{d}_r \in [0, D_r]} \left[U_r(\overline{d}_r, l) - t\overline{d}_r \right].$$

This is the *best response mapping* for user r, given the price t and the latency l. Next, we define the set-valued mapping $\mathcal{Y}(\mathbf{d})$ for $\mathbf{d}$ such that $0 \le d_r \le D_r$ as follows:

$$\mathcal{Y}_r(\mathbf{d}) = BR_r\left(t\left(\sum_s d_s\right), l\left(\sum_s d_s\right)\right).$$

We show that the mapping $\mathcal{Y}$ has a fixed point; such a fixed point will be a competitive equilibrium. First, observe that since $U_r(\cdot, l)$ is concave and continuous for fixed $l \ge 0$, the set $BR_r(t, l)$ is nonempty and convex for fixed $t, l \ge 0$, and thus $\mathcal{Y}_r(\mathbf{d})$ is nonempty and convex. To apply Kakutani's fixed point theorem [124], it remains to be shown that $\mathcal{Y}_r$ is a closed mapping.

Suppose then that we have a sequence $\mathbf{d}(n) \to \mathbf{d}$ as $n \to \infty$, such that $\mathbf{y}(n) \in \mathcal{Y}(\mathbf{d}(n))$, and $\mathbf{y}(n) \to \mathbf{y}$. We must show that $\mathbf{y} \in \mathcal{Y}(\mathbf{d})$. But this follows easily by writing the optimality conditions which define $BR_r(t, l)$. We have $\mathbf{d} \in BR_r(t, l)$ if and only if:

$$\frac{\partial^+ U_r(d_r, l)}{\partial d_r} \le t, \quad \text{if } 0 \le d_r < D_r;$$
$$\frac{\partial^- U_r(d_r, l)}{\partial d_r} \ge t, \quad \text{if } 0 < d_r \le D_r.$$

Let $l_n = l(\sum_s d_s(n))$, and let $t_n = t(\sum_s d_s(n))$; and let $l = l(\sum_s d_s)$, and let $t = t(\sum_s d_s)$. Then as $n \to \infty$, we have $l_n \to l$ and $t_n \to t$, by continuity of the functions l and t. Thus we have:

$$\frac{\partial^+ U_r(y_r(n), l_n)}{\partial d_r} - t_n \to \frac{\partial^+ U_r(y_r, l)}{\partial d_r} - t,$$

as $n \to \infty$; a similar result holds for the left directional derivative if $0 < y_r \le D_r$. (Recall we assumed U_r to have continuous directional derivatives.) From this it follows that if $\mathbf{y}(n) \in BR_r(t_n, l_n)$ for all n, then $\mathbf{y} \in BR_r(t, l)$. This establishes that $\mathcal{Y}_r$ is a closed mapping, and hence $\mathcal{Y}$ possesses a fixed point, as required. □

The previous example shows a competitive equilibrium exists; however, in general, the efficiency loss can be arbitrarily high relative to the maximal aggregate surplus, regardless of the choice of the function t. The basic intuition for this fact is well known in economics: there exists a *negative externality* between the users due to the latency l, and by using a price of the form $t(\sum_s d_s)$, the externality cannot be priced correctly. (For details on the theory of externalities, see Chapter 11 of [82].) We illustrate this first in a general setting, then specialize to a "selfish routing" example.

Example 3.6
Fix a continuous and nonnegative function $t(f)$. We will construct an example for which the competitive equilibrium has arbitrarily low efficiency relative to the maximal aggregate surplus, as follows. Suppose the system consists only of one user, with $U(d, l) = \alpha d - \beta dl$, where $\alpha > 0$ and $\beta > 0$; in addition, suppose that $l(f) = f$. Observe that in this case U satisfies Assumption 3.10, and l satisfies Assumption 3.11. We assume that the cost function C is identically zero. It is straightforward to verify that the maximal aggregate surplus in this situation is given by $\alpha^2/4\beta$, and is achieved when $d = \alpha/2\beta$.

On the other hand, suppose that d is a competitive equilibrium rate. We assume without loss of generality that $d > 0$; otherwise the efficiency loss relative to the maximal aggregate surplus is trivially 100%. Let $t = t(d)$, and $l = l(d) = d$. Since d maximizes $U(\overline{d}, l) - t\overline{d}$ over $\overline{d} \geq 0$, we must have $\alpha - \beta l - t = 0$, or $d = (\alpha - t)/\beta$. In this case the competitive equilibrium aggregate surplus is given by $\alpha d - \beta dl = t(\alpha - t)/\beta$. Thus the ratio of competitive equilibrium aggregate surplus to the maximal aggregate surplus is:

$$\frac{4t(d)(\alpha - t(d))}{\alpha^2}. \tag{3.118}$$

Thus suppose we choose a sequence of pairs (α_n, β_n), with $\alpha_n \to \infty$ as $n \to \infty$, such that the competitive equilibrium is d when the utility function is $U_n(\overline{d}, l) = \alpha_n \overline{d} - \beta_n \overline{d} l$. (Such a choice can be accomplished by choosing $\beta_n > 0$ so that $\alpha_n - \beta_n d = t$.) In this case $t(d)$ remains fixed, and the ratio (3.118) will tend to zero, as required. □

Example 3.7 (Selfish routing with elastic demands)
A related special case occurs when we set $t(f) = 0$ for all f, and $U_r(d_r, l) = u_r(d_r) - d_r l$ for all r (where $u_r(d_r)$ is a utility function satisfying Assumption 3.1). Such a model is then identical to the selfish routing model of Wardrop [144], but where users' demands are elastic (i.e., determined by utility maximization), rather than inelastic (i.e., exogenously specified); see Chapter 6 of [120] or Chapter 2 of [10]. The preceding example then can be specialized to (trivially) yield arbitrarily high efficiency loss in this setting. This is in contrast to the results of Roughgarden and Tardos [108], who show that when demands are fixed and latency functions are affine, the aggregate latency is no more than $4/3$ of the minimal aggregate latency. In our setting with elastic demands, the efficiency loss is arbitrarily high even if latency functions are affine. We note here that Chau and Sim [19] have considered a model similar to this example, and provide a bound on the efficiency loss in terms of the parameters of a given game instance; in addition, Schulz and Stier Moses have considered a related model of elastic demands and compare only the costs (not the utilities) achieved at the competitive equilibrium and the aggregate surplus maximizing allocation [114]. □

These negative results are quite trivial, and suggest that naive schemes are not sufficient to eliminate the efficiency loss due to "mispricing" that results in the context of selfish routing. In these results we have focused on competitive equilibrium, where users are price taking. When users are price taking, the particular choice of market mechanism is less essential—one could also consider a model where the strategy of each user is their total willingness-to-pay (as in Section 3.1), rather than the Cournot model of this section. The resulting mechanism has the same efficiency properties as those described here, and in particular, efficiency loss can be arbitrarily high even though users are price taking. Note that, of course, the fact that price taking behavior may lead to arbitrarily high efficiency loss suggests that price anticipating behavior may lead to arbitrarily high efficiency loss, by considering an appropriate limiting case of many users.

■ 3.7 Chapter Summary

This chapter has considered a variety of settings where the available supply of resources is elastic, as compared to the setting of the previous chapter where resources are inelastically supplied. For a game where users' strategies are the payments they are willing to make, we have shown that the efficiency loss is no more than 34% when users are price anticipating, for the setting of a single link (Theorem 3.8) as well as a network (Theorem 3.14). We also considered Cournot games, where the strategies of users are the quantities they desire. While efficiency loss is generally arbitrarily high for Cournot games when users are price anticipating, we show for some special cases that efficiency loss may be bounded (Theorems 3.20, 3.21, and 3.22). Finally, we considered some simple models for Cournot games where utility depends on latency as well as rate.

We close by comparing and contrasting the price anticipating behavior of users in the models of this chapter, and the price anticipating behavior of users discussed in Section 2.1.2. Note that for the model of that section, users need to know only the current price at the link, i.e., $\mu = \sum_r w_r/C$, as well as the capacity C of the link, to compute an optimal response. On the other hand, in the discussion of Section 3.1.2, in order to compute an optimal strategic decision users need to know not only the current price level $p(f(\mathbf{w}))$, but also the total allocated rate $f(\mathbf{w})$ and the derivative of the price $p'(f(\mathbf{w}))$ (where we have assumed for simplicity that p is differentiable). We postulate that the overhead of actually collecting such detailed information in a large scale communication network is quite high; in fact, in general users will have no knowledge of either the total allocated rate or the derivative of the price at the resource. This raises an important question of information availability when users respond to price signals: users may not react optimally, so what are users' conjectures about how their strategies affect the price? Developing more detailed models for the users' response to available price information from the network poses an interesting research direction for the future.

CHAPTER 4

Multiple Producers, Inelastic Demand

In this chapter we turn our attention to a model where multiple suppliers compete to satisfy a fixed, inelastic demand. We will discuss two motivations for such a model—the first theoretical, and the second practical. First, we will delineate the differences between "consumers" and "producers," lending justification to further theoretical investigation of models of competing firms. Next, we will relate a model consisting of multiple producers and inelastic demand to existing work on modeling electricity markets, and in particular connect our model with the extensive work on *supply function equilibrium*.

In the previous two chapters, we have considered models where multiple consumers bid for a scarce resource (either in inelastic supply, as in Chapter 2, or in elastic supply, as in Chapter 3). In each of these models, given a consumer with utility function U, we assume the net monetary payoff when the allocation is x and the payment is w is given by:

$$U(x) - w. \tag{4.1}$$

In this chapter we consider a symmetric problem, where multiple suppliers "bid" to meet a fixed demand. We describe a firm by a cost function C, which gives the monetary cost of production by the firm. If the firm produces x and receives revenue R, then the net monetary payoff to the firm is:

$$R - C(x).$$

Notice that this payoff is identical to (4.1), if we define $C(x) = -U(x)$ and $w = -R$. Thus, on the surface, one might expect the results of the previous two chapters to also carry through to games where suppliers compete, rather than consumers. However, we must be careful in defining consumers and producers. Formally, we say that an agent in a market is a *consumer* if he has nondecreasing utility (or, equivalently, nonincreasing cost) in the amount of resource allocated. Conversely, we say an agent in a market is a *producer* if he has nondecreasing cost (or, equivalently, nonincreasing utility) in the amount of resource allocated. Because of this distinction, if $U(x)$ is the utility of a consumer, then $C(x) = -U(x)$ will not be the cost of a producer. Thus, a further theoretical investigation of market-clearing mechanisms for environments with multiple competing producers and inelastic demand is required.

An important motivation for considering such models may be found in the evolution of modern markets for generation of electricity. Demand for electricity, particularly in the short run, is characterized by low elasticity with respect to price, i.e., changes in price do not lead to significant changes in the level of demand; see, e.g., [131], Section 1-7.3. We emphasize here that this characteristic is primarily an artifact of the architecture of electricity pricing as it exists today, where prices seen by consumers do not typically vary on timescales shorter than a month. Indeed, recent efforts at changing the pricing architecture for electricity demand have focused on ensuring the responsiveness of demand on shorter timescales [91].

In light of the short run price inelasticity of demand, markets for generation today operate by setting a price for electricity so that the aggregate supply offered by generators meets the demand requirements of a given region. This raises a simple market design question: given a fixed, inelastic demand, how should a market mechanism be designed to yield an efficient allocation of generation—that is, an allocation which minimizes production cost? In particular, as in the previous two chapters, we desire efficiency even if firms are price anticipating.

Before considering various models for market mechanisms, we fix some terminology. We will assume that N generators compete to satisfy a fixed demand $D > 0$, and that each generator n submits a *supply function* $S_n(p)$ to a central clearinghouse, describing the amount of electricity the generator is willing to produce at a given price per unit p. The clearinghouse then "clears the market" by choosing a price p such that $\sum_n S_n(p) = D$. For each firm, we also let $P_n(S)$ denote the inverse of the supply function, i.e., $P_n(S_n(p)) = p$. (These definitions are made informally, without regard to ensuring that market-clearing prices or inverse supply functions exist; we will consider these technical issues more carefully in the remainder of the chapter.)

We start by considering simple structures for the supply functions, the well known models of Bertrand [12] and Cournot [23] competition. In Bertrand competition, each firm n chooses a price p_n at which it is willing to supply any amount of electricity; that is, the inverse supply functions chosen by the firms are $P_n(s) = p_n$ for all $s > 0$. The market manager then chooses the lowest price among those offered by the n firms, and the entire demand is supplied at this price (possibly split among multiple firms offering the lowest price). However, as observed by Shapiro [119], Bertrand competition is characterized by the fact that equilibria may fail to exist when marginal production cost of each firm is not linear.

We next consider instead a mechanism where each firm chooses a fixed quantity it is willing to supply, regardless of the price; this is known as Cournot competition [23].

In this case each firm n chooses a quantity q_n, and $S_n(p) = q_n$ for all n. However, when the price elasticity of demand is zero, then such a mechanism is not well defined; and furthermore, if the price elasticity of demand is low, then it is straightforward to check that Cournot equilibria may have arbitrarily high efficiency loss when some firms are price anticipating [30].

Thus both Bertrand and Cournot mechanisms will not typically yield mechanisms which are efficient when firms are price anticipating, and demand is inelastic; in fact, in general these mechanisms may not even have well-defined market-clearing prices. For this reason, both theory and practice in construction of markets for electricity generation have focused on *supply function bidding*, or *supply function equilibrium* (SFE). In such markets, the strategy of each firm is not limited to one scalar (either price or quantity), but rather consists of an entire function $S_n(p)$ describing the amount of electricity a generator is willing to produce at any price p; note the relation to the demand function bidding discussed by Wilson [146] (see Section 2.1). Thus the strategy space of each firm is infinite dimensional.

Grossman [49] and Hart [54] provide concrete examples of SFE models. Grossman, in particular, suggested the investigation of equilibria in supply functions as a means to eliminate the efficiency losses due to market power in industries with only a small number of firms competing. His analysis shows that in the presence of fixed startup costs to the firms, it is possible for a supply function equilibrium to achieve full efficiency; however, in general it is difficult to guarantee that the number of supply function equilibria is small, and other inefficient supply function equilibria may exist.

The seminal work in the study of supply function equilibria is the paper of Klemperer and Meyer [69]. The authors begin by showing that, in the absence of uncertainty, nearly any production allocation can be supported as a supply function equilibrium. They then show that if demand is uncertain, then the range of equilibria is dramatically reduced; and that in equilibrium, the range of possible prices and allocations range between those achieved at Bertrand and Cournot equilibria. The key contribution of Klemperer and Meyer in this work is the identification of uncertainty in demand as a factor which reduces the number of equilibria; indeed, this insight builds on earlier work by the same authors in [68] which investigates preferences firms might have between prices as strategies (vertical supply functions) or quantities as strategies (horizontal supply functions).

The SFE work of Klemperer and Meyer sparked activity in the electricity market modeling literature, in no small part because these markets actually operate in practice by having generators submit complete supply functions (see, e.g., [136] for a description of the recent United States Federal Energy Regulatory Commission guidelines for Standard Market Design for the power industry). The first applications of SFE to electricity markets appeared in the papers of Bolle [16] and Green and Newbery [47, 48].

We now briefly survey some of the developments of this stream of literature. Our goal is not to give an exhaustive survey (for a more complete list of references, see for example [30]; and for a discussion of some of the issues involved, see [147]). Rather, we aim to give background on the existing models as a contrast to the model we build in this chapter.

The initial papers of [16] and [48] considered SFE models with continuous and differentiable supply functions submitted by suppliers. On the other hand, in practice generators submit discretized supply functions (typically step functions). As would be expected, such an assumption on the supply functions may preclude existence of Nash equilibria in pure strategies, a result shown by von der Fehr and Harbord [141]. As one response to this negative result, Supatgiat et al. consider a game where generators submit single price-quantity pairs as bids, and bids are accepted at increasing price until demand is met; the authors characterize Nash equilibria for this game [133].

Another key assumption in the original work of Klemperer and Meyer [69] is that demand varies with price (i.e., that demand is elastic). Rudkevich et al. extend the original analysis of Klemperer and Meyer to games where demand is completely inelastic, but possibly uncertain [111]. Building on this model and the conclusions of von der Fehr and Harbord in [141], Anderson and Philpott characterize supply function equilibria with inelastic demand under general assumptions on the cost functions of the individual firms, and then investigate the loss of revenue if firms must approximate both their own supply functions and the supply functions of competitors [1].

We make two observations about this line of work as it relates to our models in this chapter. First, the original model of Klemperer and Meyer in [69] required that the firms competing in a market have identical cost functions; characterizing supply function equilibria when firms are not necessarily symmetric is a much harder problem. As discussed by Baldick et al. [7], asymmetric markets are typically handled by making linearity assumptions on the structure of the supply functions submitted. Baldick and Hogan [8, 9] justify such an assumption by showing that, in general, supply function equilibria other than affine supply functions will be “unstable” (i.e., unlikely to persist in practice); however, their theoretical conclusion relies on an assumption that marginal costs are affine for all firms. In summary, therefore, the complexity of the SFE model places restrictions on the types of environments which can be successfully analyzed.

Our second observation is that the line of modeling electricity markets has focused almost entirely on using the SFE framework for its predictive power. In other words, by solving the SFE model for an appropriate set of assumptions, the papers discussed above hope to lend insight into the operation of power markets which require generators to submit complete supply schedules as bids. But because there may be a multiplicity of equilibria, an explicit understanding of efficiency losses in these games has not been developed. Papers such as the work of Rudkevich et al. [111] do suggest, however, that in the presence of inelastic demand, price anticipating behavior can lead to significant deviations from perfectly efficient allocations.

These observations lead us to consider a substantially different approach in this chapter. First, we aim to create a market-clearing mechanism where the strategy space of each firm is simple (as in Bertrand or Cournot competition), but where the description of each firm's supply function is rich enough to ensure that both competitive and Nash equilibria exist (as in SFE model). Second, we desire that such a mechanism always yields full efficiency when firms are price taking, regardless of the firms' cost functions; and we also hope that the efficiency loss remains bounded when firms are price anticipating, again regardless of the firms' cost functions (provided the number of firms is greater than two). Thus the question we pose is a mechanism design question: given that a market designer wishes to use a single price to determine the allocation of demand, what is a reasonable restricted class of supply functions for which good efficiency properties can be guaranteed? (In fact, we will see in Section 5.2 that the mechanism we construct in this chapter minimizes efficiency loss among all market-clearing mechanisms satisfying certain assumptions.)

Finally, we note here a caveat to the model of this chapter. We have adopted the stance that market-clearing mechanisms are desirable because the use of a single price to ensure supply equals demand has long been an architectural feature of modern electricity markets. However, if the goal of the electricity market is only to ensure an efficient allocation of production across the firms, even if firms are price anticipating, then a reasonable solution might be found in the traditional elements of the theory of mechanism design (see [46] and Chapter 23 of [82]). For the moment we simply note this issue requires further investigation; we defer a detailed discussion to Chapter 6.

Chapter Outline

The remainder of the chapter is organized as follows. In Section 4.1, we precisely define the market mechanism we are considering; in particular, we assume that each firm submits a supply function of the form $S(p, w) = D - w/p$, where D is the demand and w is a nonnegative scalar chosen by the firm. The market then chooses a price so that aggregate supply is equal to demand. In Section 4.1.1, we assume that firms are price taking, and show there exists a competitive equilibrium; furthermore, at this competitive equilibrium the resulting allocation minimizes aggregate production cost. In Section 4.1.2, we assume instead that firms are price anticipating, and establish existence and uniqueness of a Nash equilibrium as long as more than two firms compete. In Section 4.2 we consider the aggregate production cost at a Nash equilibrium relative to the minimal possible aggregate production cost. As long as more than two firms are competing, we show that the ratio of Nash equilibrium production cost to the minimal production cost is no worse than $1 + 1/(N - 2)$, where N is the number of firms in the market. We conclude by considering two extensions of the model, first to mitigate the possibility of negative supply by the firms (discussed in Section 4.3), and second to apply to settings with stochastic demand (in Section 4.4). The latter section demonstrates that the efficiency loss result of Section 4.2 carries over even to a setting where demand is inelastic but stochastically determined, by showing that in such an instance it is *as if* firms play a game with deterministic demand but different cost functions.

■ 4.1 Preliminaries

Suppose $N > 2$ firms compete to satisfy a known inelastic demand $D > 0$. Let s_n denote the amount produced and supplied by firm n. We assume that firm n incurs a cost $C_n(s_n)$ when it produces s_n units; we assume that cost is measured in monetary units. We make the following assumption on the cost functions C_n.

Assumption 4.1
For each n, the cost function $C_n(s_n)$ is continuous, with $C_n(s_n) = 0$ if $s_n \leq 0$. Over the domain $s_n \geq 0$, the cost function $C_n(s_n)$ is convex and strictly increasing.

We observe that the assumption of convexity is quite strong in the context of electricity markets, since startup costs may be significant for a generator, thus making the overall cost function nonconvex. However, for the purposes of analytical simplicity we consider in this chapter only the setting where cost is convex; extending this model to a situation where costs may be nonconvex remains an open problem.

Given complete knowledge and centralized control of the system, a natural problem for the market manager to try to solve is the following optimization problem:

SYSTEM:

$$\text{minimize} \quad \sum_n C_n(s_n) \tag{4.2}$$

$$\text{subject to} \quad \sum_n s_n = D; \tag{4.3}$$

$$s_n \geq 0, \quad n = 1, \ldots, N. \tag{4.4}$$

The objective function in the previous equation is the *aggregate cost*. This is the natural adaptation of the concept of aggregate surplus to a setting where demand is inelastic; see Section 1.1. Note that we can restrict the feasible region to s ≥ 0, since C_n is strictly increasing if $s_n \geq 0$ and $C_n(s_n) = 0$ if $s_n \leq 0$. Since the objective function is continuous and the feasible region is compact, an optimal solution s exists; since the feasible region is convex, if the functions C_n are strictly convex, then the optimal solution is unique.

We consider the following market mechanism for production allocation. Each firm n submits a *supply function* to the market manager, which gives (as a function of price) the amount the firm is willing to produce. We will assume the supply functions are

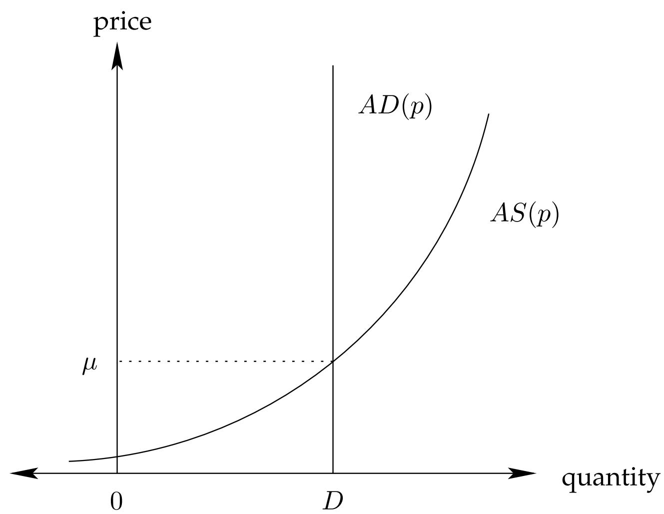

Figure 4-1. The market-clearing process with inelastic demand: Each producer n chooses a parameter w_n, which maps to the supply function $S(p, w_n) = D - w_n/p$. This defines the aggregate supply function $AS(p) = \sum_n S(p, w_n) = ND - \sum_n w_n/p$. The aggregate demand function is $AD(p) = D$ for all p. The price μ is chosen so that supply equals demand, i.e., so that $D = AD(\mu) = AS(\mu) = ND - \sum_n w_n/p$.

chosen from a parametrized family of supply functions. Formally, we assume that firm n submits a parameter $w_n \geq 0$ to the market manager. The supply function indicates that at a price $p > 0$, firm n is willing to supply $S(p, w_n)$ units given by:

$$S(p, w_n) = D - \frac{w_n}{p}. \tag{4.5}$$

We now assume that the market manager chooses the price $p(\mathbf{w}) > 0$ to *clear the market*, i.e., so that $\sum_s S(p(\mathbf{w}), w_n) = D$; see Figure 4-1. Such a choice is only possible if $\sum_n w_n > 0$, in which case:

$$p(\mathbf{w}) = \frac{\sum_n w_n}{(N-1)D}. \tag{4.6}$$

On the other hand, if $\sum_n w_n = 0$, then $S(p, w_n) = D$ for all n, regardless of the value of p; so we fix the following conventions:

$$S(0, 0) = D, \quad \text{and} \quad p(\mathbf{0}) = 0. \tag{4.7}$$

(This makes the function p continuous in $\mathbf{w}$.)

We note here that the mechanism we are considering is related to the mechanisms of Chapters 2 and 3, where users choose demand functions of the form $D(p, w_r) = w_r/p$ (recall Sections 2.1 and 3.1). The parameter w_n may be interpreted as the revenue that firm n is willing to *forego*; this may be seen since pD is the total pool of revenue when the price is p, and $pS(p, w_n) = pD - w_n$ is the revenue to firm n when the price is p. While this seems on the surface to be an unnatural bidding method for firms, we will discuss in Section 5.2 the extent to which the mechanism we consider here is uniquely defined by certain desirable properties.

Before continuing, we pause here to examine the implications of using the same mechanism as in Chapter 2 to allocate demand among the producers. In such a mechanism, each firm n submits a bid w_n; the market-clearing price is set to $\mu = \sum_n w_n/D$; and firm n produces $s_n = w_n/\mu = w_n D/(\sum_m w_m)$, while receiving revenue w_n. When firm n is price anticipating, then, the payoff to firm n is:

$$w_n - C_n\left(\frac{w_n}{\sum_m w_m} D\right).$$

But now consider a strategy where firm n lets w_n grow arbitrarily large: the revenue to firm n approaches infinity, while the cost remains bounded by $C_n(D)$. Thus we find that a Nash equilibrium does not exist for such a mechanism when firms are price anticipating. More generally, it is possible to show that such a mechanism does not ensure payoffs to firms are concave when they are price anticipating, and hence Nash equilibria cannot be guaranteed to exist; we will investigate the consequences of such a restriction further in Section 5.2 of Chapter 5. For the purposes of the current chapter, such a restriction leads us to consider the parametrized supply function class described by (4.5).

In the remainder of the section, we consider two different models for how firms might interact with the price mechanism. In Section 4.1.1, we consider a model where firms do not anticipate the effect of their bids on the price, and establish existence of a competitive equilibrium. Furthermore, this competitive equilibrium leads to an allocation which is an optimal solution to *SYSTEM*. In Section 4.1.2, we change the model and assume firms are price anticipating, and establish existence and uniqueness of a Nash equilibrium. Section 4.2 then considers the loss of efficiency at this Nash equilibrium, relative to the optimal solution to *SYSTEM*.

■ 4.1.1 Price Taking Firms and Competitive Equilibrium

In this section, we consider a *competitive equilibrium* between the firms and the market manager [82]. A central assumption in the definition of competitive equilibrium is that each firm does not anticipate the effect of its payment w_n on the price, i.e., each firm acts as a *price taker*. In this case, given a price $\mu > 0$, firm n acts to maximize the following profit function over $w_n \geq 0$:

$$P_n(w_n; \mu) = \mu S(\mu, w_n) - C_n(S(\mu, w_n)). \tag{4.8}$$

If we substitute using (4.5) and (4.6), then for $\mu > 0$ we have:

$$P_n(w_n; \mu) = (\mu D - w_n) - C_n\left(D - \frac{w_n}{\mu}\right).$$

The first term represents the revenue to firm n when the price is μ and the firm supplies $S(\mu, w_n)$ units; the second term represents the cost to the firm of producing $S(\mu, w_n)$ units. Observe that since cost is measured in monetary units, the payoff is *quasilinear* in money.

We now say a pair $(\mathbf{w}, \mu)$ where $\mathbf{w} \geq 0$ and $\mu > 0$ is a *competitive equilibrium* if firms maximize their payoff as defined in (4.8), and the market is cleared by setting the price μ according to (4.6):

$$P_n(w_n; \mu) \quad \geq \quad P_n(\overline{w}_n; \mu) \quad \text{for } \overline{w}_n \geq 0, \;\; n = 1, \ldots, N; \tag{4.9}$$

$$\mu \quad = \quad \frac{\sum_n w_n}{(N-1)D}. \tag{4.10}$$

We show that when firms are price takers, there exists a competitive equilibrium, and the resulting allocation is an optimal solution to *SYSTEM*. This is formalized in the following theorem.

Theorem 4.1
Suppose that Assumption 4.1 is satisfied and that $N > 1$. Then there exists a competitive equilibrium, i.e., a vector $\mathbf{w} = (w_1, \ldots, w_N) \geq 0$ and a scalar $\mu > 0$ satisfying (4.9)-(4.10). *In this case, the vector $\mathbf{s}$ defined by $s_n = S(\mu, w_n)$ is an optimal solution to SYSTEM.*

Proof. The key idea in the proof is to use Lagrangian techniques to establish that the equilibrium conditions (4.9)-(4.10) are identical to the optimality conditions for the problem *SYSTEM*, under the identification $s_n = S(\mu, w_n)$ for each n.

Step 1: Given $\mu > 0$, $\mathbf{w}$ satisfies (4.9) *if and only if $w_n \in [0, \mu D]$ for all n, and:*

$$\frac{\partial^- C_n(S(\mu, w_n))}{\partial s_n} \quad \leq \quad \mu, \quad \text{if } 0 \leq w_n < \mu D; \tag{4.11}$$

$$\frac{\partial^+ C_n(S(\mu, w_n))}{\partial s_n} \quad \geq \quad \mu, \quad \text{if } 0 < w_n \leq \mu D. \tag{4.12}$$

To see that these conditions are necessary and sufficient, first note that firm n would never bid more than μD when the price is μ. If $w_n > \mu D$, then $S(\mu, w_n) < 0$, so the payoff $P_n(w_n; \mu)$ becomes negative; on the other hand, $P_n(\mu D; \mu) = 0$. Thus if w_n satisfies (4.9) for firm n, then $w_n \in [0, \mu D]$. To complete the proof, we note only that convexity of C_n implies concavity of P_n; and thus w_n satisfies (4.9) if and only if $w_n \in [0, \mu D]$, and w_n satisfies the optimality conditions (4.11)-(4.12).

Step 2: There exists a vector $\mathbf{s} \geq 0$ *and a scalar* $\mu > 0$ *such that:*

$$\frac{\partial^- C_n(s_n)}{\partial s_n} \leq \mu, \quad \text{if } s_n > 0; \tag{4.13}$$

$$\frac{\partial^+ C_n(s_n)}{\partial s_n} \geq \mu, \quad \text{if } s_n \geq 0; \tag{4.14}$$

$$\sum_n s_n = D. \tag{4.15}$$

The vector $\mathbf{s}$ *is then an optimal solution to SYSTEM.* As discussed above, at least one optimal solution to *SYSTEM* exists since the feasible region is compact and the objective function is continuous. We form the Lagrangian for the problem *SYSTEM*:

$$\mathcal{L}(\mathbf{s}, \mu) = \sum_n C_n(s_n) - \mu \left(\sum_n s_n - D \right)$$

Here the second term is a penalty for the demand constraint. The Slater constraint qualification ([13], Section 5.3) holds for the problem *SYSTEM* at the point $\mathbf{s} = 0$, since then $0 = \sum_n s_n < C$; this guarantees the existence of a *Lagrange multiplier* μ. In other words, since the objective function is convex and the feasible region is convex, a feasible vector s is optimal if and only if there exists $\mu \geq 0$ such that the conditions (4.13)-(4.15) hold. Since there exists at least one optimal solution s to *SYSTEM*, there exists at least one pair $(\mathbf{s}, \mu)$ satisfying (4.13)-(4.15). We see that $\mu > 0$ from (4.13), since $s_n > 0$ for at least one firm n.

Step 3: If the pair $(\mathbf{s}, \mu)$ *satisfies* (4.13)-(4.15)*, and we let* $w_n = \mu(D - s_n)$*, then the pair* $(\mathbf{w}, \mu)$ *satisfies* (4.9)-(4.10). By Step 2, $\mu > 0$; thus, under the identification $w_n = \mu(D - s_n)$, (4.15) becomes equivalent to (4.10). Furthermore, (4.13)-(4.14) become equivalent to (4.11)-(4.12); by Step 1, this guarantees that (4.9) holds.

Step 4: Suppose $\mathbf{w}$ *and* $\mu > 0$ *satisfy* (4.9)-(4.10)*. Let* $s_n = S(\mu, w_n)$ *for each* n*. Then there exists* $\overline{\mu} > 0$ *such that the pair* $(\mathbf{s}, \overline{\mu})$ *satisfies* (4.13)-(4.15). First note that (4.10) is equivalent to (4.15) under the identification $s_n = S(\mu, w_n)$. Next, we observe that if $0 \leq s_n < D$ for all n, then $0 < w_n \leq \mu D$ for all n. Thus the conditions (4.11)-(4.12) become equivalent to the conditions (4.13)-(4.14), for the pair $(\mathbf{s}, \mu)$. Thus the claim is proven if $0 \leq s_n < D$ for all n; in this case we let $\overline{\mu} = \mu$.

On the other hand, suppose that $s_n = D$ for some n, and $s_m = 0$ for $m \neq n$; thus $w_n = 0$ and $w_m = \mu D$ for $m \neq n$. Let $\overline{\mu} = \min\{\mu, \partial^+ C_n(D)/\partial s_n\}$; note that $\overline{\mu} > 0$. Now note from (4.12), we have $\partial^+ C_m(0)/\partial s_m \geq \mu$ for $m \neq n$. Since $\overline{\mu} \leq \mu$, we conclude that (4.14) holds for $(\mathbf{s}, \overline{\mu})$. Next notice that the only firm with $s_m > 0$

is $m = n$. From (4.11), we have $\partial^- C_n(D)/\partial s_n \le \mu$; and since C_n is convex, we have $\partial^- C_n(D)/\partial s_n \le \partial^+ C_n(D)/\partial s_n$. Thus (4.13) holds for $(\mathbf{s}, \overline{\mu})$ as well, as required.

Step 5: Completing the proof. By Steps 2 and 3, there exists a vector $\mathbf{w}$ and a scalar $\mu > 0$ satisfying (4.9)-(4.10); by Step 4, the vector $\mathbf{s}$ defined by $s_n = S(\mu, w_n)$ is an optimal solution to *SYSTEM*. □

Theorem 4.1 shows that under the assumption that firms behave as price takers, there exists a strategy vector $\mathbf{w}$ where all firms have optimally chosen their bids w_n, with respect to the given price $\mu = \sum_n w_n/((N-1)D)$; and at this "equilibrium," aggregate production cost is minimized. However, when the price taking assumption is violated, the model changes into a game and the guarantee of Theorem 4.1 is no longer valid. We investigate this game in the following section.

■ 4.1.2 Price Anticipating Firms and Nash Equilibrium

We now consider an alternative model where the firms are price anticipating, rather than price takers. The key difference is that while the payoff function P_n takes the price μ as a fixed parameter in (4.8), price anticipating firms will realize that μ is set according to $\mu = p(\mathbf{w})$ from (4.6), and adjust their payoff accordingly; this makes the model a game between the N firms.

We use the notation $\mathbf{w}_{-n}$ to denote the vector of strategies of firms other than n; i.e., $\mathbf{w}_{-n} = (w_1, w_2, \ldots, w_{n-1}, w_{n+1}, \ldots, w_N)$. Given $\mathbf{w}_{-n}$, each firm n chooses w_n to maximize:

$$Q_n(w_n; \mathbf{w}_{-n}) = p(\mathbf{w}) S(p(\mathbf{w}), w_n) - C_n\left(S(p(\mathbf{w}), w_n)\right) \tag{4.16}$$

over nonnegative w_n. If we substitute for $p(\mathbf{w})$ from (4.6) and for $S(p, w_n)$ from (4.5), we have:

$$Q_n(w_n; \mathbf{w}_{-n}) = \begin{cases} \dfrac{\sum_m w_m}{N-1} - w_n - C_n\left(D - \left(\dfrac{w_n}{\sum_m w_m}\right)(N-1)D\right), & \text{if } w_n > 0; \\[2ex] \dfrac{\sum_{m \neq n} w_m}{N-1} - C_n(D), & \text{if } w_n = 0. \end{cases} \tag{4.17}$$

The payoff function Q_n is similar to the payoff function P_n, except that the firm anticipates that the network will set the price μ according to $\mu = p(\mathbf{w})$ from (4.6). A *Nash equilibrium* of the game defined by $(Q_1, \ldots, Q_N)$ is a vector $\mathbf{w} \ge 0$ such that for all n:

$$Q_n(w_n; \mathbf{w}_{-n}) \ge Q_n(\overline{w}_n; \mathbf{w}_{-n}), \quad \text{for all } \overline{w}_n \ge 0. \tag{4.18}$$

The following theorem shows that there exists a unique Nash equilibrium allocation when $N > 2$ firms compete, by showing that at a Nash equilibrium it is *as if* the firms are solving another optimization problem of the same form as the problem *SYSTEM*, but with "modified" cost functions.

Theorem 4.2
Assume that $N \geq 2$, and suppose that Assumption 4.1 is satisfied. If $N = 2$, then no Nash equilibrium exists for the game defined by $(Q_1, \ldots, Q_N)$. On the other hand, if $N > 2$, then there exists a Nash equilibrium $\mathbf{w} \geq 0$ of the game defined by $(Q_1, \ldots, Q_N)$, and it satisfies $\sum_n w_n > 0$. In this case the vector $\mathbf{s}$ defined by $s_n = S(p(\mathbf{w}), w_n)$ is the unique optimal solution to the following optimization problem:

GAME:

$$\text{minimize} \quad \sum_n \hat{C}_n(s_n) \tag{4.19}$$

$$\text{subject to} \quad \sum_n s_n = D; \tag{4.20}$$

$$s_n \geq 0, \quad n = 1, \ldots, N, \tag{4.21}$$

where

$$\hat{C}_n(s_n) = \left(1 + \frac{s_n}{(N-2)D}\right) C_n(s_n) - \frac{1}{(N-2)D} \int_0^{s_n} C_n(z)\, dz. \tag{4.22}$$

Proof. The proof proceeds in a number of steps. We first show that at a Nash equilibrium, at least two components of $\mathbf{w}$ must be positive. This suffices to show that the payoff function Q_n is concave and continuous for each firm n. We use these properties to show no Nash equilibrium exists if $N = 2$, and then restrict attention to the case $N > 2$. We then establish necessary and sufficient conditions for $\mathbf{w}$ to be a Nash equilibrium; these conditions look similar to the optimality conditions (4.11)-(4.12) in the proof of Theorem 4.1, but for "modified" cost functions defined according to (4.22). Mirroring the proof of Theorem 4.1, we then show the correspondence between these conditions and the optimality conditions for the problem *GAME*. This correspondence establishes existence of a Nash equilibrium, and uniqueness of the resulting allocation.

Step 1: If $\mathbf{w}$ is a Nash equilibrium, then at least two coordinates of $\mathbf{w}$ are positive. Fix a firm n, and suppose $w_m = 0$ for every $m \neq n$. The payoff to firm n is then:

$$Q_n(w_n; \mathbf{w}_{-n}) = \begin{cases} -C_n(D), & \text{if } w_n = 0; \\ -\dfrac{(N-2)w_n}{N-1}, & \text{if } w_n > 0. \end{cases}$$

The first expression follows by noting that when $w_n = 0$ (so that $\mathbf{w} = 0$), we have $p(\mathbf{w}) = 0$, while $S(p(\mathbf{w}), w_n) = D$ for firm n. (Recall the convention (4.7) that $S(0, 0) =$

D for all $p \geq 0$.) For the second expression, note that when $w_n > 0$, we have the inequality $S(p(\mathbf{w}), w_n) = D - (N-1)D < 0$, so $C(S(p(\mathbf{w}), w_n)) = 0$. We now see that when $w_n = 0$, firm n can profitably deviate by increasing w_n infinitesimally (since $C_n(D) > 0$); on the other hand, when $w_n > 0$, firm n can profitably deviate by infinitesimally decreasing w_n. Thus no Nash equilibrium exists where $\sum_{m \neq n} w_m = 0$. Since this holds for every firm n, we conclude at least two coordinates of $\mathbf{w}$ must be positive.

Step 2: If the vector $\mathbf{w} \geq 0$ *has at least two positive components, then the function* $Q_n(\overline{w}_n; \mathbf{w}_{-n})$ *is concave and continuous in* $\overline{w}_n$*, for* $\overline{w}_n \geq 0$*.* This follows from (4.16). When $\sum_{m \neq n} w_m > 0$, from (4.17) we have:

$$Q_n(\overline{w}_n; \mathbf{w}_{-n}) = \frac{\sum_{m \neq n} w_m}{N-1} - \frac{(N-2)\overline{w}_n}{N-1} - C_n\left(D - \left(\frac{\overline{w}_n}{\overline{w}_n + \sum_{m \neq n} w_m}\right)(N-1)D\right).$$

Indeed, when $\sum_{m \neq n} w_m > 0$, the function $\overline{w}_n / (\overline{w}_n + \sum_{s \neq r} w_s)$ is a strictly concave function of $\overline{w}_n$ (for $\overline{w}_n \geq 0$). Since C_n was assumed to be convex and nondecreasing (and hence continuous), it follows that $Q_n(\overline{w}_n; \mathbf{w}_{-n})$ is concave and continuous in $\overline{w}_n$ for $\overline{w}_n \geq 0$.

Step 3: If $N = 2$*, then no Nash equilibrium exists.* Suppose that (w_1, w_2) is a Nash equilibrium. Then by Step 1, $w_1 > 0$ and $w_2 > 0$; and by Step 2, in this case the payoff to firm 1 as a function of $\overline{w}_1 \geq 0$ is:

$$Q_1(\overline{w}_1; w_2) = w_2 - C_n\left(D - \frac{\overline{w}_1}{\overline{w}_1 + w_2} D\right).$$

The preceding expression is strictly increasing in $\overline{w}_1$, so (w_1, w_2) could not have been a Nash equilibrium. Thus no Nash equilibrium exists if $N = 2$.

Based on the preceding step, for the remainder of the proof, we will assume that $N > 2$.

Step 4: The vector $\mathbf{w}$ *is a Nash equilibrium if and only if at least two components of* $\mathbf{w}$ *are positive, and for each* n*,* $w_n \in [0, (\sum_{m \neq n} w_m)/(N-2)]$ *and the following conditions hold:*

$$\frac{\partial C_n^-(S(p(\mathbf{w}), w_n))}{\partial s_n}\left(1 + \frac{S(p(\mathbf{w}), w_n)}{(N-2)D}\right) \leq p(\mathbf{w}), \text{ if } 0 \leq w_n < \frac{\sum_{m \neq n} w_m}{N-2}; \quad (4.23)$$

$$\frac{\partial C_n^+(S(p(\mathbf{w}), w_n))}{\partial s_n}\left(1 + \frac{S(p(\mathbf{w}), w_n)}{(N-2)D}\right) \geq p(\mathbf{w}), \text{ if } 0 < w_n \leq \frac{\sum_{m \neq n} w_m}{N-2}. \quad (4.24)$$

Let $\mathbf{w}$ be a Nash equilibrium. By Steps 1 and 2, $\mathbf{w}$ has at least two positive components and $Q_n(\overline{w}_n; \mathbf{w}_{-n})$ is concave and continuous for $\overline{w}_n \geq 0$. We first observe that we must have $w_n \leq (\sum_{m \neq n} w_m)/(N-2)$; if not, then $S(p(\mathbf{w}), w_n) < 0$, and by arguing as in Step 1 of the proof of Theorem 4.1 we can show that firm n can profitably deviate by choosing $\overline{w}_n = (\sum_{m \neq n} w_m)/(N-2)$. Thus w_n must maximize $Q_n(\overline{w}_n; \mathbf{w}_{-n})$ over $0 \leq \overline{w}_n \leq (\sum_{m \neq n} w_m)/(N-2)$, and satisfy the following first order optimality conditions:

$$\begin{aligned} \frac{\partial^+ Q_n(w_n; \mathbf{w}_{-n})}{\partial w_n} &\leq 0, \quad \text{if } 0 \leq w_n < \frac{\sum_{m \neq n} w_m}{N-2}; \\ \frac{\partial^- Q_n(w_n; \mathbf{w}_{-n})}{\partial w_n} &\geq 0, \quad \text{if } 0 < w_n \leq \frac{\sum_{m \neq n} w_m}{N-2}. \end{aligned}$$

Recalling the expression for $p(\mathbf{w})$ given in (4.6), after multiplying through by $p(\mathbf{w})$ the preceding optimality conditions become:

$$\frac{\partial C_n^-(S(p(\mathbf{w}), w_n))}{\partial s_n}\left(1 - \frac{w_n}{\sum_m w_m}\right) \leq \frac{(N-2)p(\mathbf{w})}{N-1}, \quad \text{if } 0 \leq w_n < \frac{\sum_{m \neq n} w_m}{N-2}; \tag{4.25}$$

$$\frac{\partial C_n^+(S(p(\mathbf{w}), w_n))}{\partial s_n}\left(1 - \frac{w_n}{\sum_m w_m}\right) \geq \frac{(N-2)p(\mathbf{w})}{N-1}, \quad \text{if } 0 < w_n \leq \frac{\sum_{m \neq n} w_m}{N-2}. \tag{4.26}$$

We now note that by definition, we have:

$$\frac{w_n}{\sum_m w_m} = \frac{D - S(p(\mathbf{w}), w_n)}{(N-1)D}.$$

Substituting into (4.25)-(4.26) and simplifying yields (4.23)-(4.24).

Conversely, suppose that $\mathbf{w}$ has at least two strictly positive components, that $0 \leq w_n \leq (\sum_{m \neq n} w_m)/(N-2)$, and $\mathbf{w}$ satisfies (4.23)-(4.24). Then we may simply reverse the argument: by Step 2, $Q_n(\overline{w}_n; \mathbf{w}_{-n})$ is concave and continuous in $\overline{w}_n \geq 0$, and in this case the conditions (4.23)-(4.24) imply that w_n maximizes $Q_n(\overline{w}_n; \mathbf{w}_{-n})$ over $0 \leq \overline{w}_n \leq (\sum_{m \neq n} w_m)/(N-2)$. Since we have already shown that choosing $\overline{w}_n > (\sum_{m \neq n} w_m)/(N-2)$ is never optimal for firm n, we conclude $\mathbf{w}$ is a Nash equilibrium.

If we let $\mu = p(\mathbf{w})$, note that the conditions (4.23)-(4.24) have the same form as the optimality conditions (4.11)-(4.12), but for a different cost function given by $\hat{C}_n$. We now use this relationship to complete the proof in a manner similar to the proof of Theorem 4.1.

Step 5: The function $\hat{C}_n(s_n)$ defined in (4.22) *is continuous, and strictly convex and strictly increasing over $s_n \geq 0$, with $\hat{C}_n(s_n) = 0$ for $s_n \leq 0$.* Since $C_n(s_n) = 0$ for $s_n \leq 0$, it follows that $\hat{C}_n(s_n) = 0$ for $s_n \leq 0$. For $s_n \geq 0$, we simply compute the directional derivatives of $\hat{C}_n$:

$$\frac{\partial^+ \hat{C}_n(s_n)}{\partial s_n} = \left(1 + \frac{s_n}{(N-2)D}\right) \frac{\partial^+ C_n(s_n)}{\partial s_n};$$
$$\frac{\partial^- \hat{C}_n(s_n)}{\partial s_n} = \left(1 + \frac{s_n}{(N-2)D}\right) \frac{\partial^- C_n(s_n)}{\partial s_n}.$$

Since C_n is strictly increasing and convex, for $0 \leq s_n < \overline{s}_n$ we will have:

$$0 \leq \frac{\partial^+ \hat{C}_n(s_n)}{\partial s_n} < \frac{\partial^- \hat{C}_n(\overline{s}_n)}{\partial s_n} \leq \frac{\partial^+ \hat{C}_n(\overline{s}_n)}{\partial s_n}.$$

This guarantees that $\hat{C}_n$ is strictly increasing and strictly convex over $s_n \geq 0$.

Step 6: There exists a unique vector $\mathbf{s} \geq 0$ and at least one scalar $\rho > 0$ such that:

$$\left(1 + \frac{s_n}{(N-2)D}\right) \frac{\partial^- C_n(s_n)}{\partial s_n} \leq \rho, \quad \text{if } s_n > 0; \tag{4.27}$$
$$\left(1 + \frac{s_n}{(N-2)D}\right) \frac{\partial^+ C_n(s_n)}{\partial s_n} \geq \rho, \quad \text{if } s_n \geq 0; \tag{4.28}$$
$$\sum_n s_n = D. \tag{4.29}$$

The vector $\mathbf{s}$ is then the unique optimal solution to GAME. By Step 5, since $\hat{C}_n$ is continuous and strictly convex over the convex, compact feasible region for each n, we know that *GAME* has a unique optimal solution $\mathbf{s}$. As in the proof of Theorem 4.1, the Slater constraint qualification holds for *GAME*, so there exists a Lagrange multiplier ρ such that $(\mathbf{s}, \rho)$ satisfy the stationarity conditions (4.27)-(4.28), together with the constraint (4.29). The fact that $\rho > 0$ follows from (4.27), since at least one s_n is positive.

Step 7: If $\mathbf{s} \geq 0$ and $\rho > 0$ satisfy (4.27)-(4.29)*, then the vector $\mathbf{w}$ defined by $w_n = (D - s_n)\rho$ is a Nash equilibrium.* First, observe by this definition that $w_n \geq 0$, from (4.29) and the fact that $s_n \geq 0$ for all n. Furthermore, since $s_n \geq 0$, of course we have $(1 + 1/(N-2))s_n \geq 0$; it is thus straightforward to check that we have:

$$w_n = (D - s_n)\rho \leq \frac{((N-2)D + s_n)\,\rho}{N-2} = \frac{\sum_{m \neq n} w_m}{N-2}.$$

Finally, at least two components of $\mathbf{w}$ are strictly positive, since otherwise we have $s_{n_1} = s_{n_2} = D$ for some $n_1 \neq n_2$, in which case $\sum_n s_n > D$, which contradicts (4.29).

By Step 4, to check that $\mathbf{w}$ is a Nash equilibrium, we must only check the stationarity conditions (4.23)-(4.24). We simply note that under the identification $w_n = (D - s_n)\rho$, using (4.29) we have that:

$$\rho = \frac{\sum_n w_n}{(N-1)D} = p(\mathbf{w}); \text{ and } s_n = D - \frac{w_n}{\rho} = S(\rho, w_n).$$

Substitution of these expressions into (4.27)-(4.28) leads immediately to (4.23)-(4.24). Thus $\mathbf{w}$ is a Nash equilibrium.

Step 8: If $\mathbf{w}$ *is a Nash equilibrium, then there exists a scalar* $\rho > 0$ *such that the vector* $\mathbf{s}$ *defined by* $s_n = S(p(\mathbf{w}), w_n)$ *satisfies* (4.27)-(4.29). We simply reverse the argument of Step 7. Since $\mathbf{w}$ is a Nash equilibrium, by Step 1 $\sum_n w_n > 0$, so $p(\mathbf{w}) > 0$; thus $\sum_n s_n = D$, i.e., (4.29) is satisfied. By Step 4, $\mathbf{w}$ satisfies (4.23)-(4.24). We now consider two possibilities. First suppose that $0 \leq s_n < D$ for all n; then let $\rho = p(\mathbf{w})$. In this case $\rho > 0$ and $0 < w_n \leq (\sum_{m \neq n} w_m)/(N-2)$ for all n, so (4.23)-(4.24) become equivalent to (4.27)-(4.28). On the other hand, suppose that $s_n = D$ for some n; then $s_m = 0$ for all $m \neq n$. We define ρ by:

$$\rho = \min\left\{ p(\mathbf{w}), \left(1 + \frac{1}{N-2}\right) \frac{\partial^+ C_n(D)}{\partial s_n} \right\}.$$

Note that again, we have $\rho > 0$. We now argue as in Step 4 of the proof of Theorem 4.1. By combining (4.24) with the definition of ρ, we see that (4.28) is satisfied. Finally, since $s_m > 0$ only for $m = n$, we combine (4.23) with the fact that $\partial^- C_n(s_n) \leq \partial^+ C_n(s_n)$ (by convexity) to see that (4.27) holds, as required.

Step 9: There exists a Nash equilibrium $\mathbf{w}$*, and the vector* $\mathbf{s}$ *defined by* $s_n = S(p(\mathbf{w}), w_n)$ *is the unique optimal solution of GAME.* This conclusion is now straightforward. Existence follows by Steps 6 and 7. Uniqueness of the resulting production vector $\mathbf{s}$, and the fact that $\mathbf{s}$ is an optimal solution to *GAME*, follows by Steps 6 and 8. □

An interesting parallel between the proof of Theorem 4.2 and the proof of Theorem 2.2 is evident: both theorems use "modified" objective functions for *SYSTEM*, and show that solving the resulting optimization problem yields the allocation achieved at a Nash equilibrium when market participants are price anticipating. In the case of Theorem 2.2, the modified utility functions $\hat{U}_r$ are strictly concave; and in the case of Theorem 4.2, the modified cost functions $\hat{C}_n$ are strictly convex. In each case, these properties ensure uniqueness of the resulting allocation at a Nash equilibrium.

4.2 Efficiency Loss

We let $\mathbf{s}^S$ denote an optimal solution to *SYSTEM*, and let $\mathbf{s}^G$ denote the unique optimal solution to *GAME*. We now investigate the efficiency loss of this system: that is, to what extent does the aggregate production cost increase because firms try to game the market? To answer this question, we must compare the cost $\sum_n C_n(s_n^G)$ obtained when the firms fully evaluate the effect of their actions on the price, and the cost $\sum_n C_n(s_n^S)$ obtained by choosing the allocation which minimizes aggregate cost. (We know, of course, that $\sum_n C_n(s_n^G) \geq \sum_n C_n(s_n^S)$ by definition of $\mathbf{s}^S$.) The following theorem shows that we may explicitly bound the efficiency loss.

Theorem 4.3
Assume that $N > 2$, and suppose that Assumption 4.1 is satisfied. If $\mathbf{s}^S$ is any optimal solution to SYSTEM, and $\mathbf{s}^G$ is the unique optimal solution to GAME, then:

$$\sum_n C_n(s_n^G) \leq \left(1 + \frac{1}{N-2}\right) \sum_n C_n(s_n^S). \tag{4.30}$$

Furthermore, this bound is tight: for every $\varepsilon > 0$ and $N > 2$, there exists a choice of cost functions C_n, $n = 1, \ldots, N$, such that:

$$\sum_n C_n(s_n^G) \geq \left(1 + \frac{1}{N-2} - \varepsilon\right) \sum_n C_n(s_n^S). \tag{4.31}$$

Proof. We exploit the structure of the modified cost functions $\hat{C}_n$ to prove the result. Let $G_n(s_n) = \int_0^{s_n} C_n(z)\, dz$. Then by our assumptions on C_n, G_n is a convex, continuous, nondecreasing function for $s_n \geq 0$, with $G_n(0) = 0$. This implies that for $s_n \geq 0$ we have $s_n C_n(s_n) - G_n(s_n) \geq 0$; from the definition of $\hat{C}_n$ in (4.22), we conclude that $\hat{C}_n(s_n) \geq C_n(s_n)$ for $s_n \geq 0$. This yields:

$$\sum_n \hat{C}_n(s_n^G) \geq \sum_n C_n(s_n^G). \tag{4.32}$$

On the other hand, notice that for $s_n \geq 0$, we have $G_n(s_n) \geq 0$. Thus for $0 \leq s_n \leq D$, we have:

$$\hat{C}_n(s_n) \leq C_n(s_n) + \left(\frac{s_n}{(N-2)D}\right) C_n(s_n) \leq \left(1 + \frac{1}{N-2}\right) C_n(s_n).$$

This yields:

$$\sum_n \hat{C}_n(s_n^S) \leq \left(1 + \frac{1}{N-2}\right) \sum_n C_n(s_n^S). \tag{4.33}$$

Since $\mathbf{s}^G$ is an optimal solution to *GAME*, we know that $\sum_n \hat{C}_n(s_n^G) \leq \sum_n \hat{C}_n(s_n^S)$. Combining this inequality with (4.32) and (4.33) yields the bound (4.30).

It remains to be shown that the bound is tight, i.e., that (4.31) holds; we prove this via an example. We fix $D > 0$, and assume we are given $N > 2$. Choose t such that $D/N < t < D$, and choose δ such that $0 < \delta < 1$. Consider the following cost functions:

$$C_1(s_1) = \begin{cases} \delta s_1, & \text{if } s_1 \leq t; \\ s_1 - t + \delta t, & \text{if } s_1 \geq t; \end{cases}$$

$$C_n(s_n) = \alpha s_n, \quad n = 2, \ldots, N,$$

where

$$\alpha = \frac{1 + \dfrac{t}{(N-2)D}}{1 + \dfrac{D-t}{(N-1)(N-2)D}}.$$

Thus C_1 is piecewise linear, and C_n is linear for $n = 2, \ldots, N$. It is straightforward to check that $t > D/N$ implies $\alpha > 1$; thus, the unique optimal solution to *SYSTEM* is given by $s_1^S = D$, $s_n^S = 0$ for $n = 2, \ldots, N$, and we have $\sum_n C_n(s_n^S) = D - t + \delta t$.

Let $s_1 = t$, and $s_n = (D-t)/(N-1)$ for $n = 2, \ldots, N$. We claim that $\mathbf{s}$ is the unique optimal solution to *GAME*. To see this, let $\rho = 1 + t/((N-2)D)$. Then we have:

$$\left(1 + \frac{s_1}{(N-2)D}\right) \frac{\partial^- C_1(s_1)}{\partial s_1} = \delta \left(1 + \frac{t}{(N-2)D}\right) \leq \rho;$$

$$\left(1 + \frac{s_n}{(N-2)D}\right) \frac{\partial^+ C_n(s_n)}{\partial s_n} = 1 + \frac{t}{(N-2)D} = \rho;$$

$$\left(1 + \frac{s_n}{(N-2)D}\right) \frac{\partial C_n(s_n)}{\partial s_n} = 1 + \frac{t}{(N-2)D} = \rho, \quad n = 2, \ldots, N;$$

$$\sum_n s_n = D.$$

These conditions are identical to (4.27)-(4.29), so we conclude $\mathbf{s}$ is the unique optimal solution to *GAME*. Observe that:

$$\sum_n C_n(s_n) = \delta t + \left(\frac{1 + \dfrac{t}{(N-2)D}}{1 + \dfrac{D-t}{(N-1)(N-2)D}} \right) (D - t).$$

Thus we have:

$$\frac{\sum_n C_n(s_n)}{\sum_n C_n(s_n^S)} = \left(\frac{1}{\delta t + (D-t)}\right)\left(\delta t + \left(\frac{1 + \dfrac{t}{(N-2)D}}{1 + \dfrac{D-t}{(N-1)(N-2)D}}\right)(D-t)\right).$$

Now let $t \to D$ and $D \to 1$, while $D/N < t < D$, and let $\delta \to 0$ so that $\delta t/(D-t) \to 0$, e.g., let $\delta = (D-t)^2$. Then the preceding ratio converges to $1+1/(N-2)$, as required.□

The preceding theorem shows that in the worst case, aggregate cost rises by no more than a factor $1 + 1/(N-2)$ when firms are able to anticipate the effects of their actions on the price of the resource. Furthermore, this bound is essentially tight. In comparing Theorem 4.3 with the previous results of this thesis, particularly Theorem 2.6, we make two key observations. First, the bound obtained by using the modified cost functions $\hat{C}_n$ is in fact tight in the case of Theorem 4.3. As discussed in Section 2.2, however, the upper bound on efficiency loss derived by using the modified utility functions $\hat{U}_r$ defined in (2.19) is 50%, which is not tight (the efficiency loss is 25% in the worst case, as shown in Theorem 2.6). Second, we note that in the case of Theorem 4.3, the efficiency loss $1/(N-2)$ approaches zero as the number of firms N grows large, even though the firms are price anticipating. This is a form of a competitive limit theorem [82]; however, note that this result holds even if only a small number of firms continue to remain dominant as $N \to \infty$ (i.e., we do not require any symmetry constraints on the cost functions of the firms). In the model of Section 2.2, such a limit only held under the assumption that all users consumed a negligible fraction of the resource in the limit (see Corollary 2.8). Indeed, in an industry with one large firm and many small firms, we do not expect to achieve full efficiency; nevertheless, the mechanism described in this chapter ensures this is the case.

■ 4.3 Negative Supply

One undesirable feature of the parametrized supply functions we have chosen is that they do not ensure each firm will have nonnegative supply at the market-clearing price. While we have shown that the supply of each firm is nonnegative at both the competitive equilibrium and at the Nash equilibrium, nonequilibrium bidding may lead to negative supply to some firms. In this section we consider a simple modification to the basic model which corrects this issue.

We continue to assume that $S(p, w_n)$ and the market-clearing price $p(\mathbf{w})$ are defined as before. We fix a *minimum liability* $W > 0$, such that no firm will ever have to pay more than W when the market is cleared. Thus, if $p(\mathbf{w})S(p(\mathbf{w}, w_n) < -W$, then firm n

only pays W to the market manager. Formally, the payoff of firm n now becomes:

$$\overline{Q}_n(w_n; \mathbf{w}_{-n}) = \max\{-W, p(\mathbf{w})S(p(\mathbf{w}), w_n)\} - C_n(S(p(\mathbf{w}), w_n)). \tag{4.34}$$

One interpretation of this game is as follows. Each firm submits a participation fee of W to the market manager. The game is then played as before, and the market manager clears the market. At the resulting allocation, any required payment higher than W by a firm is forgiven.

We have the following proposition.

Proposition 4.4
Assume that $N > 2$, and suppose that Assumption 4.1 is satisfied. Then $\mathbf{w}$ is a Nash equilibrium of the game defined by $(Q_1, \ldots, Q_N)$ if and only if $\mathbf{w}$ is a Nash equilibrium of the game defined by $(\overline{Q}_1, \ldots, \overline{Q}_N)$.

Proof. The proof technique uses the fact that any firm n can always guarantee itself $Q_n(w_n; \mathbf{w}_{-n}) > -W$, given the value of $\mathbf{w}_{-n}$. To see this, first suppose that $\sum_{m \neq n} w_m > 0$. Then if $w_n = (\sum_{m\neq n} w_m)/(N-2)$, we will have $Q_n(w_n; \mathbf{w}_{-n}) = 0$ (as shown in Step 1 of the proof of Theorem 4.2). On the other hand, if $\sum_{m\neq n} w_m = 0$, then for sufficiently small $w_n > 0$, we will have $Q_n(w_n; \mathbf{w}_{-n}) > -W$; this also follows by Step 1 of the proof of Theorem 4.2.

Thus, suppose that $\mathbf{w}$ is a Nash equilibrium of the game defined by $(Q_1, \ldots, Q_N)$; then we have $Q_n(w_n; \mathbf{w}_{-n}) > -W$. If $\mathbf{w}$ is not a Nash equilibrium for the game defined by $(\overline{Q}_1, \ldots, \overline{Q}_N)$, then there exists a firm n and $\overline{w}_n \geq 0$ such that $\overline{Q}_n(\overline{w}_n; \mathbf{w}_{-n}) > \overline{Q}_n(w_n; \mathbf{w}_{-n})$. It follows that $\overline{Q}_n(\overline{w}_n; \mathbf{w}_{-n}) > -W$, so that $\overline{w}_n$ is a profitable deviation for firm n in the game defined by $(Q_1, \ldots, Q_N)$—a contradiction. Thus $\mathbf{w}$ is a Nash equilibrium of the game defined by $(\overline{Q}_1, \ldots, \overline{Q}_N)$.

Conversely, suppose that the vector $\mathbf{w}$ is a Nash equilibrium of the game defined by $(\overline{Q}_1, \ldots, \overline{Q}_N)$. Then we must have $\overline{Q}_n(w_n; \mathbf{w}_{-n}) > -W$ for all n, so that $\overline{Q}_n(w_n; \mathbf{w}_{-n}) = Q_n(w_n; \mathbf{w}_{-n})$. An argument similar to the preceding paragraph then shows that $\mathbf{w}$ is a Nash equilibrium of the game defined by $(Q_1, \ldots, Q_N)$. □

While this extension to the game is appealing from a market implementation point of view, we must be careful in interpreting the preceding result. Suppose that $\mathbf{w}$ is a composite strategy vector where $p(\mathbf{w})S(p(\mathbf{w}), w_n) < -W$; in particular, $S(p(\mathbf{w}), w_n) < 0$. In this case we will have $\sum_{m\neq n} S(p(\mathbf{w}), w_m) > D$—that is, the remaining firms will be producing *excess supply*. In an economy with free disposal, this does not pose any problem. However, in the context of electricity markets, such a situation indicates a misalignment of supply and demand, and can induce instability in the power grid. In general, then, addressing the possibility of negative supply when $\mathbf{w}$ is not a Nash equilibrium remains an important implementation-dependent issue.

4.4 Stochastic Demand

In this section we consider a model where demand is stochastic, rather than predetermined; the model developed is analogous to the one in Section 2.5.1. Suppose that the demand is D, where $0 \le D \le D_{\max}$, with distribution $\mathbb{P}$ (thus $\mathbb{P}(D > D_{\max}) = 0$). We assume that $\mathbb{E}[D] = \int_0^{D_{\max}} D d\mathbb{P}(D) > 0$. We also assume, as in Section 4.1, that N firms compete for the demand.

We will define an allocation in terms of the *fractions* allocated to each firm, rather than the absolute amount of resource allocated. Formally, we define the problem *SYSTEM* as:

SYSTEM:

$$\text{minimize} \quad \sum_n \mathbb{E}[C_n(\pi_n D)] \tag{4.35}$$

$$\text{subject to} \quad \sum_n \pi_n = 1; \tag{4.36}$$

$$\pi_n \ge 0, \quad n = 1, \ldots, N. \tag{4.37}$$

Notice that this problem chooses the fractions π_n allocated to each resource optimally *ex ante*; that is, before the true supply has been realized.

As in Section 2.5.1, our key insight in analyzing this model is that stochastic demand is equivalent to a model with deterministic demand $D = 1$, for an appropriate choice of cost functions. Formally, for each firm n, define $\overline{C}_n$ as follows:

$$\overline{C}_n(\pi_n) = \mathbb{E}[C_n(\pi_n D)]. \tag{4.38}$$

We have the following proposition.

Proposition 4.5
Suppose that Assumption 4.1 is satisfied by the cost functions $C_1, \ldots, C_N$. Then Assumption 4.1 is also satisfied by the cost functions $\overline{C}_1, \ldots, \overline{C}_N$.

Proof. Observe that $\overline{C}_n(\pi_n) = 0$ for $\pi_n \le 0$, since C_n satisfies Assumption 4.1. We next show that $\overline{C}_n$ is continuous, using the same argument as in Proposition 2.18. Suppose that $\pi_n^k \to \pi_n$ as $k \to \infty$. Then $C_n(\pi_n^k D) \to C_n(\pi_n D)$ as $k \to \infty$ for $0 \le D \le D_{\max}$. Since $C_n(s_n)$ is nonnegative and nondecreasing, there exists $\varepsilon > 0$ such that for sufficiently large k we have $0 \le C_n(\pi_n^k D) \le C_n(\pi_n D_{\max} + \varepsilon)$. Thus we can apply the bounded convergence theorem to conclude that as $k \to \infty$, $\mathbb{E}[C_n(\pi_n^k D)] \to \mathbb{E}[C_n(\pi_n D)]$. Thus $\overline{C}_n$ is continuous.

It remains to be shown that $\overline{C}_n(\pi_n)$ is convex and strictly increasing for $\pi_n \ge 0$.

First fix $\delta > 0$, and $\pi_n^1, \pi_n^2 \geq 0$. Then for fixed $D > 0$ we have:

$$C_n(\delta\pi_n^1 D + (1-\delta)\pi_n^2 D) \leq \delta C_n(\pi_n^1 D) + (1-\delta)C_n(\pi_n^2 D).$$

Taking expectations shows that $\overline{C}_n(\pi_n)$ is convex for $\pi_n \geq 0$. Finally, suppose that $\pi_n^1 > \pi_n^2 \geq 0$. Since C_n is strictly increasing, for $D > 0$ we have $C_n(\pi_n^1 D) > C_n(\pi_n^2 D)$. Taking expectations shows that $\overline{C}_n(\pi_n)$ is strictly increasing for $\pi_n \geq 0$ (since $\mathbb{E}[D] > 0$). □

The preceding lemma allows us to extend the main results of Sections 4.1 and 4.2 to the setting of stochastic demand. We start with the following proposition.

Proposition 4.6
Suppose that Assumption 4.1 is satisfied. Then there exists a vector $\boldsymbol{\pi}^S$ that is an optimal solution to (4.35)-(4.37). *Furthermore, $\boldsymbol{\pi}^S$ is an optimal solution to* (4.35)-(4.37) *if and only if $\boldsymbol{\pi}^S$ is an optimal solution to* (4.2)-(4.4) *with cost functions $\overline{C}_1, \ldots, \overline{C}_N$ and demand $D = 1$.*

Proof. From Proposition 4.5, the objective function (4.35) is continuous and the feasible region (4.36)-(4.37) is compact. Furthermore, under the identification (4.38), the problem (4.35)-(4.37) becomes equivalent to (4.2)-(4.4). □

We use an analogue of the pricing mechanism developed in Section 4.1. First, each firm n chooses a parameter w_n. Next, the demand D is realized. The market manager then takes as input the supply function $S(p, w_n) = D - w_n/p$ for each firm n, and clears the market by choosing $p(\mathbf{w})$ according to (4.6). Note that when a firm n chooses a parameter w_n, it is still as if the firm has chosen the supply function $D - w_n/p$; but now the supply function depends on the eventual realization of the demand D. The payoff to each firm is then the *expected* profit. Since we will focus on price anticipating firms in this section, we redefine their payoff explicitly in terms of the strategy vector $\mathbf{w}$. Formally, by substituting using (4.5) and (4.6), the payoff to firm n is given by:

$$\overline{Q}_n(w_n; \mathbf{w}_{-n}) = \begin{cases} \mathbb{E}\left[\dfrac{\sum_m w_m}{N-1} - w_n - C_n\left(D - \left(\dfrac{w_n}{\sum_m w_m}\right)(N-1)D\right)\right], & \text{if } w_n > 0; \\[2ex] \mathbb{E}\left[\dfrac{\sum_{m \neq n} w_m}{N-1} - C_n(D)\right], & \text{if } w_n = 0. \end{cases} \tag{4.39}$$

If we substitute $\overline{C}_n(\pi_n) = \mathbb{E}[C_n(\pi_n D)]$ (cf. (4.38)), then we have:

$$\overline{Q}_n(w_n; \mathbf{w}_{-n}) = \begin{cases} \dfrac{\sum_m w_m}{N-1} - w_n - \overline{C}_n\left(1 - \left(\dfrac{w_n}{\sum_m w_m}\right)(N-1)\right), & \text{if } w_n > 0; \\ \\ \dfrac{\sum_{m \neq n} w_m}{N-1} - \overline{C}_n(1), & \text{if } w_n = 0. \end{cases}$$

Now observe that $\overline{Q}_n(w_n; \mathbf{w}_{-n})$ is identical to the payoff $Q_n(w_n; \mathbf{w}_{-n})$ if we substitute the cost function $\overline{C}_n$ and demand $D = 1$ in the definition (4.17). This observation leads to the following proposition.

Proposition 4.7
Assume that $N > 2$, and suppose that Assumption 4.1 is satisfied. Then there exists a Nash equilibrium $\mathbf{w} \geq 0$ for the game defined by $(\overline{Q}_1, \dots, \overline{Q}_N)$. Furthermore, $\mathbf{w} \geq 0$ is a Nash equilibrium for the game defined by $(\overline{Q}_1, \dots, \overline{Q}_N)$ if and only if $\mathbf{w}$ is a Nash equilibrium for the game defined by $(Q_1, \dots, Q_N)$ when the cost function of each firm n is $\overline{C}_n$ and the demand is $D = 1$.

The next proposition shows that the aggregate cost at a Nash equilibrium is no worse than a factor $1 + 1/(N-2)$ larger than the aggregate cost at an optimal solution to *SYSTEM*. The intuition is clear: we simply apply Theorem 4.3 to a game where the cost function of each firm n is $\overline{C}_n$, and the demand is $D = 1$.

Theorem 4.8
Suppose that $N > 2$ and that Assumption 4.1 is satisfied. Let $\mathbf{w}$ be a Nash equilibrium of the game defined by $(\overline{Q}_1, \dots, \overline{Q}_N)$, and define:

$$\pi_n^G = 1 - \left(\frac{w_n}{\sum_m w_m}\right)(N-1).$$

Then π_n^G is the fraction of demand produced by firm n at the Nash equilibrium $\mathbf{w}$. Furthermore, if $\boldsymbol{\pi}^S$ is any optimal solution to SYSTEM, then:

$$\sum_n \mathbb{E}[C_n(\pi_n^G D)] \leq \left(1 + \frac{1}{N-2}\right) \sum_n \mathbb{E}[C_n(\pi_n^S D)].$$

Proof. We know $\overline{C}_1, \dots, \overline{C}_N$ satisfy Assumption 4.1 (from Proposition 4.5). Thus by applying Propositions 4.6 and 4.7 together with Theorem 4.3, the result follows. □

Note that while in Section 2.5.1 the capacity C is a random variable with support in $[0, \infty)$, in this section we require the demand D to be a random variable with compact support $[0, D_{\max}]$. In fact, it is clear from the proof of Proposition 4.5 that the key requirement is that $C_n(\pi_n D)$ must be integrable for $\pi_n \geq 0$. But this is only possible if D has bounded support; otherwise, by choosing a cost function C_n which approaches infinity rapidly enough, we can guarantee that $\mathbb{E}[C_n(\pi_n D)] = \infty$. Thus if we want to ensure that $\overline{C}_n(\pi_n)$ is finite whenever C_n satisfies Assumption 4.1, the random variable D must have bounded support.

■ 4.5 Chapter Summary

In this chapter we have considered a model of multiple producers competing to satisfy a fixed demand. We have shown that if we restrict firms to bidding supply functions of the form $S(p, w) = D - w/p$, then there exist fully efficient competitive equilibria when firms are price taking (Theorem 4.1), and Nash equilibria when $N > 2$ firms are price anticipating (Theorem 4.2). Furthermore, the Nash equilibrium aggregate cost is no worse than a factor $1 + 1/(N-2)$ higher than the minimal aggregate cost (Theorem 4.3). Finally, all these results continue to hold even in the case that the demand is stochastic.

An important difference between this chapter and Chapter 2 is that in considering "worst case" games, we cannot simply replace the cost functions C_n by linear approximations, as we did in the proof of Theorem 2.6. The reason is that if we replace the function $C_n(s_n)$ by $\tilde{C}_n(s_n) = C_n(\overline{s}_n) + C_n'(\overline{s}_n)(s_n - \overline{s}_n)$, where $\overline{s}_n > 0$, the resulting function $\tilde{C}_n(s_n)$ is not always nonnegative. Thus we must consider instead the cost function $[\tilde{C}_n(s_n)]^+$. This difficulty—that we cannot replace general cost functions with linear cost functions—precludes an immediate extension of the results of this chapter to a network context, as in the general network game of Section 2.5.2.

Developing an extension of this game to a network context thus remains an important issue. Such an extension is also complicated by the complex physical properties of an electrical network—electricity cannot be stored, and does not "flow" through the network according to the same laws governing either Internet traffic or road transportation networks. Instead, the equilibrium electric flow in the power grid requires solution of a global physical optimization problem, and thus changes at a local node of the grid can potentially have wide reaching consequences [115]. For this reason the material of this chapter represents only the starting point in a broader examination of power network market structure.

In moving from Chapter 2 to 3, we moved from a model with inelastic supply to one with inelastic demand. A similar undertaking may be done for the model of this chapter, by changing from the assumption of inelastic demand to assuming that demand is in fact elastic and varies with price. However, the same supply function mechanism cannot be considered directly, as the inelastic demand D is required as a parameter to the supply functions, which take the form $S(p, w) = D - w/p$. Thus an important research direction for the future involves developing a supply function bidding mechanism for elastic demand which can ensure bounded efficiency loss even when firms are price anticipating.

We close by noting that, of course, the scheme we consider is not very natural; it is difficult to imagine that electricity firms will subscribe to a model which requires them to submit as a parameter the total amount of revenue they are willing to forego (cf. Section 4.1). However, the key contribution of this chapter is a counterpoint to the extensive work on supply function equilibria: that by restricting the strategy spaces of the firms intelligently, it may be possible to prevent arbitrarily high efficiency losses. A second motivation for the scheme of Section 4.1 will be presented in the following chapter, where we will show that it is the scheme which minimizes efficiency loss among a class of market mechanisms satisfying certain desirable axioms (see Section 5.2).

CHAPTER 5

Characterization Theorems

This thesis has considered the design of market mechanisms which can ensure reasonably efficient allocation of resources. In the first three chapters of the thesis, we considered specific mechanisms, and analyzed their efficiency properties both when market participants are price taking, and when market participants are price anticipating. This chapter focuses on designing market mechanisms more generally, for the settings of either multiple consumers and inelastic supply (Chapter 2) or multiple producers and inelastic demand (Chapter 4).

In this chapter we ask an axiomatic question: are the mechanisms we have chosen "desirable" among a class of mechanisms satisfying certain "reasonable" properties? Of course, an answer to this question hinges on our definition of "desirable" and "reasonable." Defining desirability is the simpler of the two tasks: we consider a mechanism to be desirable if it achieves a fully efficient allocation when users are price taking, and if it minimizes efficiency loss when users are price anticipating. Importantly, we ask for these efficiency properties *independent* of the characteristics of the market participants (i.e., their cost functions or utility functions). That is, the mechanisms we seek are those that perform well under broad assumptions on the nature of the preferences of market participants.

The mechanisms we are trying to characterize must minimize efficiency loss among the class of "reasonable" mechanisms. We are thus led to define mathematically the conditions we would like a market mechanism to satisfy. The first such conditions are straightforward: we would like to consider only mechanisms where market participants submit either demand or supply functions, and where the resulting allocation is chosen by fixing a single price to clear the market. We require that this market-clearing price be unique, and that the mechanism is "smooth" with respect to the strategies of the market participants (in a sense we make precise later). We refer to such mechanisms as *smooth market-clearing mechanisms*.

An important additional condition we impose is that the strategy space of each market participant should be "simple," which we interpret as *low dimensional*. Formally, we will focus on mechanisms for which the strategy space of each market participant is $\mathbb{R}^+$; that is, each market participant chooses a scalar, which is a parameter that determines either his demand or supply function as input to the market-clearing mechanism. The motivation here is twofold. First, as in the models of Chapter 2, in a communication network setting information flow is limited; and in particular, we would like to implement a market with as little overhead as possible. Thus keeping the strategy spaces of the users low dimensional is a first reasonable goal. A second motivation comes from the models of Chapter 4, where we discussed the possibility that general supply function equilibria may lead to arbitrarily low efficiency at a Nash equilibrium. Thus we might hope that by restricting the strategy space of the firms, we can reduce this inefficiency; indeed, this intuition is borne out by the result of Section 4.2.

We then impose two mathematical constraints on the class of mechanisms we consider. First, we require that payoffs of market participants are concave when they are price taking; and second, we require their payoffs to be concave when they are price anticipating. The former requirement eases characterization of competitive equilibria, while the latter eases characterization of Nash equilibria. The requirement of these conditions is certainly a debatable point; indeed, in the course of this chapter we will discuss the extent to which these conditions are necessary, and when they might be relaxed.

Within this class of mechanisms, we characterize exactly the mechanisms which achieve full efficiency at competitive equilibria and minimize efficiency loss at Nash equilibria. We will find that when multiple consumers compete for a resource in inelastic supply, the unique mechanism satisfying these conditions is the mechanism studied in Section 2.1. Similarly, we will find that when multiple producers compete to satisfy an inelastic demand, the unique mechanism satisfying these conditions is the mechanism studied in Section 4.1.

Chapter Outline

The remainder of the chapter is organized as follows. In Section 5.1, we consider the setting of multiple consumers competing for a resource in inelastic supply. We start in Section 5.1.1 by proving that the mechanism considered in Section 2.1 minimizes efficiency loss among all smooth market-clearing mechanisms for which users' payoffs are concave both when they are price taking or price anticipating; where the demand functions chosen by the users are nonnegative; and where full efficiency is achieved if users are price taking. We also show that for any such mechanism, there exists a unique Nash equilibrium when users are price anticipating. We then proceed in Section 5.1.2 to show that we can remove the requirement that users' payoffs must be concave when they are price taking, if we require our mechanisms to be well defined for any value of the inelastic supply. Finally, in Section 5.1.3, we present a specialized mechanism which can guarantee arbitrarily low efficiency loss for the case of two users. This mechanism sets only a single common price for both users, but cannot be defined in terms of the users submitting scalar parametrized demand functions.

In Section 5.2, we turn our attention to the setting of multiple producers competing to satisfy an inelastic demand. We prove that the mechanism considered in Section 4.1 minimizes efficiency loss among all smooth market-clearing mechanisms for which firms' payoffs are concave both when they are price taking or price anticipating; where the supply functions chosen by the firms are uniformly no larger than the demand; and where full efficiency is achieved if firms are price taking. We also compare this result with the results of Section 5.1.

■ 5.1 Multiple Consumers, Inelastic Supply

In this section we will consider the same resource allocation model as in Section 2.1. We will ask: what is the mechanism which minimizes efficiency loss when users are price anticipating, in a class of mechanisms with certain desirable properties?

Formally, we consider a collection of users bidding to receive a share of a finite, infinitely divisible resource of capacity C. We begin by describing the users. Each user has a utility function $U : \mathbb{R}^+ \to \mathbb{R}^+$ (where $\mathbb{R}^+ = [0, \infty)$). We assume that U is continuous, strictly increasing, and concave on $[0, \infty)$. We also assume that U is continuously differentiable on $(0, \infty)$, with finite right directional derivative at 0, denoted $U'(0)$. Note that these conditions are identical to Assumption 2.1. Let $\mathcal{U}$ denote the set of possible utility functions; i.e.:

$$\mathcal{U} = \big\{U : \mathbb{R}^+ \to \mathbb{R}^+ \mid U \text{ is continuous, strictly increasing, concave on } [0,\infty), \\ \text{and continuously differentiable on } [0,\infty), \text{with } U'(0) < \infty\big\}.$$

Note that although we make rather strong differentiability assumptions, these are not essential to the argument; however, they ease the technical presentation. We let R denote the number of users, and let $\mathbf{U} = (U_1, \dots, U_R)$ denote the vector of utility functions, where U_r is the utility function of user r. We call a pair $(R, \mathbf{U})$, where $R > 1$ and $\mathbf{U} \in \mathcal{U}^R$, a *utility system*; our goal will be to design a resource allocation mechanism which is efficient for all utility systems.

We assume that utility is measured in monetary units; thus, if user r receives a rate allocation d_r, but must pay w_r, he receives a net payoff given by:

$$U_r(d_r) - w_r.$$

Given any vector of utility functions $\mathbf{U} \in \mathcal{U}^R$, our goal is to maximize aggregate utility, as defined in the following problem:

SYSTEM$(C, R, \mathbf{U})$:

$$\text{maximize} \quad \sum_{r=1}^{R} U_r(d_r) \tag{5.1}$$

$$\text{subject to} \quad \sum_{r=1}^{R} d_r \leq C; \tag{5.2}$$

$$\mathbf{d} \geq 0. \tag{5.3}$$

We will say that $\mathbf{d}$ *solves SYSTEM*$(C, R, \mathbf{U})$ if $\mathbf{d}$ is an optimal solution to (5.1)-(5.3), given the utility system $(R, \mathbf{U})$.

In general, the utility system $(R, \mathbf{U})$ is unknown to the mechanism designer, so a mechanism must be designed to elicit information from the users. We will consider two different possibilities: first, that the capacity $C > 0$ is fixed before the mechanism is chosen; and alternatively, that the chosen mechanism must be well defined for all positive values of the capacity C. The former case is discussed in the following section, and the latter case is discussed in Section 5.1.2.

5.1.1 A First Characterization Theorem

In this section, we will consider smooth market-clearing mechanisms given a fixed capacity $C > 0$. We have the following definition.

Definition 5.1
Given $C > 0$, a smooth market-clearing mechanism for C *is a differentiable function $D : (0, \infty) \times [0, \infty) \to \mathbb{R}^+$ such that for all R, and for all nonzero $\boldsymbol{\theta} \in (\mathbb{R}^+)^R$, there exists a unique solution $p > 0$ to the following equation:*

$$\sum_{r=1}^{R} D(p, \theta_r) = C.$$

We let $p_D(\boldsymbol{\theta})$ denote this solution.

Note that while this definition implicitly restricts the strategy θ_r of each user to $\mathbb{R}^+$, this fact is inessential; the subsequent analysis can be adapted to hold even if the strategy space of each user is allowed to be $[c, \infty)$, where $c \in \mathbb{R}$. We also note that the market-clearing price is undefined if $\boldsymbol{\theta} = \mathbf{0}$. As we will see below, when we formulate a game between consumers for a given mechanism D, we will assume that the payoff to all players is $-\infty$ if the composite strategy vector is $\boldsymbol{\theta} = \mathbf{0}$. Note that this is slightly different from the definition in Section 2.1, where the payoff is $U(0)$ to a player with utility function U who submits a strategy $\theta = 0$. We will discuss this distinction further later; we simply note for the moment that it does not affect the results of this section.

Our definition of smooth market-clearing mechanism generalizes the demand function interpretation of the mechanism discussed in Section 2.1. We recall that in that development, each user submits a demand function of the form $D(p, \theta) = \theta/p$, and the link manager chooses a price $p_D(\boldsymbol{\theta})$ to ensure that $\sum_{r=1}^{R} D(p, \theta_r) = C$. Thus, for this mechanism, we have $p_D(\boldsymbol{\theta}) = \sum_{r=1}^{R} \theta_r / C$ if $\boldsymbol{\theta} \neq 0$. Another related example is provided by $D(p, \theta) = \theta/\sqrt{p}$; in this case it is straightforward to verify that $p_D(\boldsymbol{\theta}) = (\sum_{r=1}^{R} \theta_r / C)^2$.

We will restrict attention to a particular class of smooth market-clearing mechanisms for C denoted $\mathcal{D}(C)$, which we define as follows.

Definition 5.2
Given $C > 0$, the class $\mathcal{D}(C)$ consists of all smooth market-clearing mechanisms D for C such that the following conditions are satisfied:

1. *For all $U \in \mathcal{U}$, a user's payoff is concave if he is price taking; that is, for all $p > 0$ the function:*
$$U(D(p, \theta)) - pD(p, \theta)$$
is concave for $\theta \geq 0$.

2. *For all $U_r \in \mathcal{U}$, a user's payoff is concave if he is price anticipating; that is, for all R, and for all $\boldsymbol{\theta}_{-r} \in (\mathbb{R}^+)^R$, the function:*
$$U_r(D(p_D(\boldsymbol{\theta}), \theta_r)) - p_D(\boldsymbol{\theta}) D(p_D(\boldsymbol{\theta}), \theta_r)$$
is concave in $\theta_r > 0$ if $\boldsymbol{\theta}_{-r} = \mathbf{0}$, and concave in $\theta_r \geq 0$ if $\boldsymbol{\theta}_{-r} \neq \mathbf{0}$.

3. *The demand functions are nonnegative; i.e., for all $p > 0$ and $\theta \geq 0$, $D(p, \theta) \geq 0$.*

We pause here to briefly discuss the three conditions in the previous definition. The first two conditions ease characterization of equilibria in terms of only first order conditions. The first condition allows us to characterize competitive equilibria in terms of only first order conditions, as we did in the proof of Theorem 2.1. The second condition allows us to characterize Nash equilibria in terms of only first order conditions, a property we exploited in the proof of Theorem 2.2; indeed, some such assumption is generally used to guarantee existence of pure strategy Nash equilibria [104]. Finally, the third condition is a normalization condition, which ensures that regardless of the bid of a user, he is never required to *supply* some quantity of the resource (which would be the case if we allowed $D(p, \theta) < 0$).

Of these conditions, the least desirable one is the first, that users' payoffs are concave when they are price taking. First, we note that concavity of users' payoffs is not necessarily required to ensure existence of a competitive equilibrium. Indeed, as long as we assume that for any $p > 0$, the range of $D(p, \theta)$ spans the entire interval $[0, \infty)$

for $\theta \geq 0$, then we can consider the following convex problem for each user r:

$$\max_{d_r \geq 0} U_r(d_r) - pd_r.$$

After determining an optimal d_r, user r only needs to choose θ_r such that $D(p, \theta_r) = d_r$. In this way we could establish existence of competitive equilibria for settings where Condition 1 in Definition 5.2 does not necessarily hold. However, it is certainly the case that concavity of payoffs eases *characterization* of competitive equilibria in terms of first order optimality conditions.

Nevertheless, an additional objection to Condition 1 in Definition 5.2 is that we expect it to apply as a *limit* of Condition 2: that is, with a "large" number of users, we might expect concavity of payoffs when users are price anticipating to lead to concavity of payoffs when users are price taking. For this reason, developing a characterization theorem which does not rely on Condition 1 of Definition 5.2 is desirable; we will turn our attention to this problem in Section 5.1.2.

In order to state the main result of this section, we must define *competitive equilibrium* and *Nash equilibrium*. Given a utility system $(R, \mathbf{U})$ and a smooth market-clearing mechanism $D \in \mathcal{D}(C)$, we say that a nonzero vector $\boldsymbol{\theta} \in (\mathbb{R}^+)^R$ is a competitive equilibrium if, for $\mu = p_D(\boldsymbol{\theta})$, there holds for all r:

$$\theta_r \in \arg\max_{\overline{\theta}_r \geq 0} \left[U_r(D(\mu, \overline{\theta}_r)) - \mu D(\mu, \overline{\theta}_r) \right]. \tag{5.4}$$

Similarly, we say that a nonzero vector $\boldsymbol{\theta} \in (\mathbb{R}^+)^R$ is a Nash equilibrium if there holds for all r:

$$\theta_r \in \arg\max_{\overline{\theta}_r \geq 0} Q_r(\overline{\theta}_r; \boldsymbol{\theta}_{-r}). \tag{5.5}$$

where

$$Q_r(\theta_r; \boldsymbol{\theta}_{-r}) = \begin{cases} U_r(D(p_D(\boldsymbol{\theta}), \theta_r)) - p_D(\boldsymbol{\theta}) D(p_D(\boldsymbol{\theta}), \theta_r), & \text{if } \boldsymbol{\theta} \neq \mathbf{0}; \\ -\infty, & \text{if } \boldsymbol{\theta} = \mathbf{0}. \end{cases} \tag{5.6}$$

Notice that the payoff is $-\infty$ if the composite strategy vector is $\boldsymbol{\theta} = \mathbf{0}$, since in this case no market-clearing price exists.

Our interest is in the worst-case ratio of aggregate utility at any Nash equilibrium to the optimal value of *SYSTEM*$(C, R, \mathbf{U})$ (termed the the "price of anarchy" by Pa-

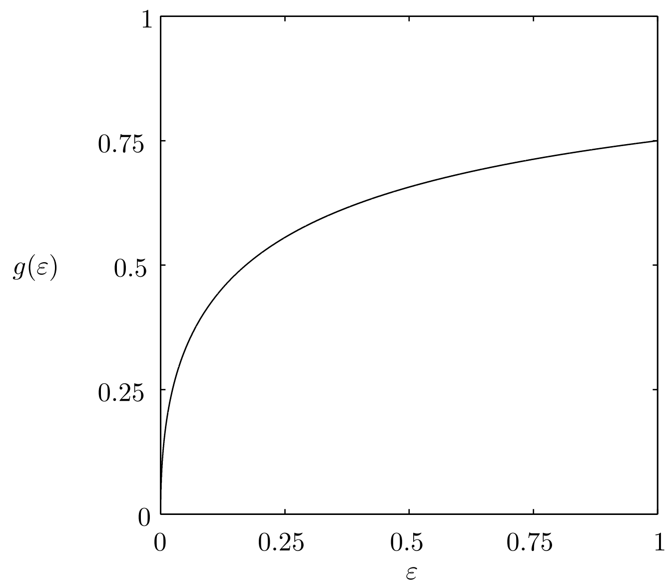

Figure 5-1. The function $g(\varepsilon)$ in Theorem 5.1: The function $g(\varepsilon)$ is defined for $0 \leq \varepsilon \leq 1$ in (5.7). Note that $g(\varepsilon)$ is strictly increasing, with $g(0) = 0$ and $g(1) = 3/4$.

padimitriou [99]). Formally, for $D \in \mathcal{D}(C)$ we define a constant $\rho(C, D)$ as follows:

$$\rho(C, D) = \inf \left\{ \frac{\sum_{r=1}^{R} U_r(D(p_D(\boldsymbol{\theta}), \theta_r))}{\sum_{r=1}^{R} U_r(d_r)} \middle| R > 1, \mathbf{U} \in \mathcal{U}^R, \mathbf{d} \text{ solves } \mathit{SYSTEM}(C, R, \mathbf{U}) \text{ and } \boldsymbol{\theta} \text{ is a Nash equilibrium} \right\}$$

Note that since all $U \in \mathcal{U}$ are strictly increasing and nonnegative, and $C > 0$, the aggregate utility $\sum_{r=1}^{R} U_r(d_r)$ is strictly positive for any utility system $(R, \mathbf{U})$ and any optimal solution $\mathbf{d}$ to *SYSTEM*$(C, R, \mathbf{U})$. However, Nash equilibria may not exist for some utility systems $(R, \mathbf{U})$; in this case we set $\rho(C, D) = -\infty$.

The following theorem shows that among smooth market-clearing mechanisms for C for which there always exists a fully efficient competitive equilibrium, the mechanism proposed in Section 2.1 minimizes efficiency loss when users are price anticipating.

Theorem 5.1
Given $C > 0$, let $D \in \mathcal{D}(C)$ be a smooth market-clearing mechanism for C such that for all utility systems $(R, \mathbf{U})$, there exists a competitive equilibrium $\boldsymbol{\theta}$ such that $(D(p_D(\boldsymbol{\theta}), \theta_r), r = 1, \ldots, R)$ solves SYSTEM$(C, R, \mathbf{U})$. Then:

1. *There exists a concave, strictly increasing, differentiable, and invertible function* $B : (0, \infty) \to (0, \infty)$ *such that for all* $p > 0$ *and* $\theta \geq 0$*:*

$$D(p, \theta) = \frac{\theta}{B(p)}.$$

2. *For any utility system* $(R, \mathbf{U})$*, there exists a unique Nash equilibrium.*

3. *Define the function* $g(\varepsilon)$ *for* $0 \leq \varepsilon \leq 1$ *according to:*

$$g(\varepsilon) = \begin{cases} \dfrac{3}{4}, & \text{if } \varepsilon = 1; \\ \\ \dfrac{2\sqrt{\varepsilon} - 3\varepsilon + \varepsilon^2}{(1-\varepsilon)^2}, & \text{if } 0 \leq \varepsilon < 1. \end{cases} \tag{5.7}$$

Then g *is continuous and strictly increasing over* $0 \leq \varepsilon \leq 1$*, with* $g(0) = 0$*; see Figure 5-1. Furthermore, if we define:*

$$\varepsilon^* = \inf_{p > 0} \left[\frac{pB'(p)}{B(p)} \right], \tag{5.8}$$

then $0 \leq \varepsilon^* \leq 1$*, and:*

$$\rho(C, D) = g(\varepsilon^*).$$

In particular, $\rho(C, D) \leq 3/4$*, and this bound is met with equality if and only if* $D(p, \theta) = \Delta\theta/p$ *for some* $\Delta > 0$*.*

Proof. The proof proceeds as follows. We first use Condition 1 in Definition 5.2 to show that any mechanism $D \in \mathcal{D}(C)$ must be of the form $D(p, \theta) = a(p) + b(p)\theta$. We then show that $a(p) = 0$, and $b(p) > 0$; thus $D(p, \theta) = b(p)\theta$. Finally, we explicitly determine conditions that must be satisfied by $B(p) = 1/b(p)$, and compute the worst case efficiency loss for any mechanism satisfying these conditions.

We begin with the following lemma.

Lemma 5.2 *Let* D *be a smooth market-clearing mechanism for* $C > 0$*. Then* $D \in \mathcal{D}(C)$ *if and only if the following three properties hold:*

1. *There exist functions* $a, b : (0, \infty) \to \mathbb{R}^+$ *such that for all* $p > 0$ *and* $\theta \geq 0$*,* $D(p, \theta) = a(p) + b(p)\theta$*.*

2. *For all* $R > 1$ *and* $\boldsymbol{\theta}_{-r} \in (\mathbb{R}^+)^R$*, the functions* $D(p_D(\boldsymbol{\theta}), \theta_r)$ *and* $-p_D(\boldsymbol{\theta})D(p_D(\boldsymbol{\theta}), \theta_r)$ *are concave in* $\theta_r > 0$ *if* $\boldsymbol{\theta}_{-r} = \mathbf{0}$*, and concave in* $\theta_r \geq 0$ *if* $\boldsymbol{\theta}_{-r} \neq \mathbf{0}$*.*

Proof of Lemma. Fix $p > 0$; we show that $D(p, \cdot)$ must be both concave and convex as a function of θ, thus implying it is linear in θ. We use Condition 1 in Definition 5.2. Let

$U(d) = \alpha d$, where $\alpha > 0$; then $U \in \mathcal{U}$. Thus for $\theta \geq 0$, $(\alpha - p)D(p, \theta)$ must be concave in θ. Thus if $\alpha < p$, then $D(p, \theta)$ must be convex in θ; and if $\alpha > p$, then $D(p, \theta)$ must be concave in θ. This implies $D(p, \theta)$ is linear in θ, so there exist $a(p)$ and $b(p)$ such that $D(p, \theta) = a(p) + b(p)\theta$. Since $D(p, \theta) \geq 0$ for all $p > 0$ and $\theta \geq 0$ (from Condition 3 in Definition 5.2), by considering the case $\theta = 0$ and the limit $\theta \to \infty$ we conclude that $a(p), b(p) \geq 0$.

The second claim of the lemma again follows by considering a utility function of the form $U_r(d_r) = \alpha d_r$, and applying Condition 2 in Definition 5.2. We must have $P(\boldsymbol{\theta}) = (\alpha - p_D(\boldsymbol{\theta}))D(p_D(\boldsymbol{\theta}), \theta_r)$ concave in θ_r over the appropriate domain. Now if $D(p_D(\boldsymbol{\theta}), \theta_r)$ is not concave, then for sufficiently large α, P will not be concave; and similarly, if $p_D(\boldsymbol{\theta})D(p_D(\boldsymbol{\theta}), \theta_r)$ is not convex, then for sufficiently small α, P will not be convex. This proves that D has the two properties claimed.

To prove the reverse implication, recall that the composition of any concave strictly increasing function with a concave function is concave. Since any $U \in \mathcal{U}$ is strictly increasing and concave, we conclude that the first claim of the lemma ensures Condition 1 in Definition 5.2 is satisfied; similarly, the second claim of the lemma ensures Condition 2 is satisfied. Finally, since $a(p), b(p) \geq 0$ for all $p > 0$, the first claim of the lemma ensures that Condition 3 in Definition 5.2 is satisfied. Thus if D satisfies the three claims of the lemma, then $D \in \mathcal{D}(C)$. □

We next show that the function $a(p)$ given in the previous lemma must be identically zero, under the conditions of the theorem. To see this, suppose $a(p) > 0$ for some $p > 0$. Choose R such that $C/R < a(p)$, and choose constants $d_1, \ldots, d_R$ such that $0 < d_r < a(p)$; $\sum_{r=1}^{R} d_r = C$; and $d_r \neq d_s$ for $r \neq s$. Finally, choose strictly concave utility functions $U_1, \ldots, U_R \in \mathcal{U}$ such that $U_r'(d_r) = p$ for all r. It is then straightforward to check that $\mathbf{d}$ must be the unique optimal solution to *SYSTEM*$(C, R, \mathbf{U})$.

So now let $\boldsymbol{\theta}$ be a competitive equilibrium such that the resulting allocation solves *SYSTEM*$(C, R, \mathbf{U})$; then we must have $D(p_D(\boldsymbol{\theta}), \theta_r) = d_r$ for all r. Since $d_r \neq d_s$ for $r \neq s$, it follows that $b(p_D(\boldsymbol{\theta})) > 0$, and $\theta_r > 0$ for at least one user r. Differentiating (5.4) with $\mu = p_D(\boldsymbol{\theta})$, we must have:

$$U_r'(D(\mu, \theta_r)) \cdot \frac{\partial D(\mu, \theta_r)}{\partial \theta_r} = \mu \cdot \frac{\partial D(\mu, \theta_r)}{\partial \theta_r}.$$

Now note that $\partial D(\mu, \theta_r)/\partial \theta_r = b(\mu) > 0$, so we have:

$$U_r'(d_r) = U_r'(D(\mu, \theta_r)) = \mu = p_D(\boldsymbol{\theta}).$$

Since $U_r'(d_r) = p$ and U_r is strictly concave, this implies $p_D(\boldsymbol{\theta}) = p$. But then $a(p) + b(p)\theta_r = d_r < a(p)$, so we must have $\theta_r < 0$, which is impossible (since the strategy space of all users is assumed to $\mathbb{R}^+$). Thus any D satisfying the conditions of the theorem must have $a(p) = 0$ for all p.

We now have $D(p, \theta) = b(p)\theta$, for all $p > 0$ and $\theta \geq 0$. We note that this immediately implies $b(p) > 0$ for all p. Otherwise, arguing as in the previous paragraph, we can choose a utility system $(R, \mathbf{U})$ with a unique optimal solution $\mathbf{d}$ to $SYSTEM(C, R, \mathbf{U})$, such that $U_r'(d_r) = p$ for all r. But such an allocation can never be a competitive equilibrium if $D(p, \theta_r) = 0$ for all $\theta_r \geq 0$.

Since $b(p) > 0$, we let $B(p) = 1/b(p)$ for all $p > 0$. In addition, since $p_D(\boldsymbol{\theta})$ satisfies $\sum_{r=1}^R D(p_D(\boldsymbol{\theta}), \theta_r) = C$ for nonzero $\boldsymbol{\theta}$, we have:

$$B(p_D(\boldsymbol{\theta})) = \frac{\sum_{r=1}^R \theta_r}{C}. \tag{5.9}$$

Note that this implies:

$$D(p_D(\boldsymbol{\theta}), \theta_r) = \frac{\theta_r}{\sum_{s=1}^R \theta_s} C, \tag{5.10}$$

for nonzero $\boldsymbol{\theta}$; thus $D(p_D(\boldsymbol{\theta}), \theta_r)$ is trivially concave in θ_r.

We immediately see that B must be invertible on $(0, \infty)$; it is clearly onto, as the right hand side of (5.9) can take any value in $(0, \infty)$. Furthermore, if $B(p_1) = B(p_2) = \gamma$ for some prices $p_1, p_2 > 0$, then choosing $\boldsymbol{\theta}$ such that $\sum_{r=1}^R \theta_r / C = \gamma$, we find that $p_D(\boldsymbol{\theta})$ is not uniquely defined. Thus B is one-to-one as well, and hence invertible. Finally, note that since D is differentiable, B must be differentiable as well.

We let Φ denote the differentiable inverse of B. We will show that Φ is strictly increasing and convex. To see this, note that for nonzero $\boldsymbol{\theta}$ we have:

$$p_D(\boldsymbol{\theta}) = \Phi\left(\frac{\sum_{r=1}^R \theta_r}{C}\right).$$

We now apply the second claim of Lemma 5.2. It must be the case that $w_r(\boldsymbol{\theta}) = p_D(\boldsymbol{\theta}) D(p_D(\boldsymbol{\theta}), \theta_r)$ is convex in θ_r, for nonzero $\boldsymbol{\theta}$. Thus we must have:

$$w_r(\boldsymbol{\theta}) = \Phi\left(\frac{\sum_{s=1}^R \theta_s}{C}\right)\left(\frac{\theta_r}{\sum_{s=1}^R \theta_s} C\right) \tag{5.11}$$

convex in θ_r for nonzero $\boldsymbol{\theta}$. It is straightforward to see that Φ must be convex; if not, then by considering strategy vectors $\boldsymbol{\theta}$ for which $\boldsymbol{\theta}_{-r} = \mathbf{0}$, we can show that w_r is not convex in θ_r. It remains to be shown that Φ is strictly increasing. Since Φ is invertible, it must be monotonic; and thus Φ is either strictly increasing or strictly decreasing. We differentiate w_r with respect to θ_r:

$$\frac{\partial w_r}{\partial \theta_r}(\boldsymbol{\theta}) = \Phi'\left(\frac{\sum_{s=1}^R \theta_s}{C}\right)\left(\frac{\theta_r}{\sum_{s=1}^R \theta_s}\right) + \Phi\left(\frac{\sum_{s=1}^R \theta_s}{C}\right)\left(\frac{\sum_{s\neq r}\theta_s}{\left(\sum_{s=1}^R \theta_s\right)^2}C\right).$$

Now choose nonzero $\boldsymbol{\theta}$ such that $\theta_r = 0$, and consider infinitesimally increasing θ_r. If Φ is strictly decreasing, then the first term does not increase, and the second strictly decreases; thus the derivative of w_r falls. But this is impossible, since w_r is convex in θ_r. We conclude that Φ must be strictly increasing.

We summarize these observations in the following lemma, which also establishes the first claim of the theorem.

Lemma 5.3 *If a smooth market-clearing mechanism $D \in \mathcal{D}(C)$ satisfies the conditions of the theorem, then there exists a concave, strictly increasing, and differentiable function B : $(0,\infty) \to (0,\infty)$ such that $D(p,\theta) = \theta/B(p)$ for all $p > 0$ and $\theta \geq 0$. Furthermore, B is invertible, so that $B(p) \to 0$ as $p \to 0$ and $B(p) \to \infty$ as $p \to \infty$. Finally, given R, for nonzero $\boldsymbol{\theta} \in (\mathbb{R}^+)^R$ there holds:*

$$p_D(\boldsymbol{\theta}) = \Phi\left(\frac{\sum_{r=1}^R \theta_r}{C}\right),$$

where Φ is the inverse of B.

Conversely, if there exists such a function $B(p)$ with $D(p,\theta) = \theta/B(p)$ for all $\theta \geq 0$ and $p > 0$, and B is twice differentiable, then $D \in \mathcal{D}(C)$.

Proof of Lemma. Since B has already been shown to be invertible, and Φ has been shown to be strictly increasing and convex, it is clear that B is strictly increasing and concave, with $B(p) \to 0$ as $p \to 0$, and $B(p) \to \infty$ as $p \to \infty$.

It remains to be checked that $D \in \mathcal{D}(C)$ for any mechanism of the form $D(p,\theta) = \theta/B(p)$, where B is twice differentiable and satisfies the conditions of the lemma. It is clear that D satisfies the first and third claims in Lemma 5.2. To check the second claim, we note from (5.10) that $D(p_D(\boldsymbol{\theta}), \theta_r)$ is concave in θ_r for nonzero $\boldsymbol{\theta}$. Thus it remains to be shown that $w_r(\boldsymbol{\theta}) = p_D(\boldsymbol{\theta})D(p_D(\boldsymbol{\theta}), \theta_r)$ is convex in θ_r for nonzero $\boldsymbol{\theta}$. We twice differentiate $w_r(\boldsymbol{\theta})$ given in (5.11). Letting $\mu = \sum_{s=1}^R \theta_s/C$, we have for nonzero $\boldsymbol{\theta}$:

$$\frac{\partial^2 w_r}{\partial \theta_r^2}(\boldsymbol{\theta}) = \Phi''(\mu)\frac{\theta_r}{C^2\mu} + \frac{2\sum_{s\neq r}\theta_s}{C^2\mu^3}\left(\mu\Phi'(\mu) - \Phi(\mu)\right).$$

Since Φ is convex, the first term is nonnegative; and since $\Phi(x) \to 0$ as $x \to 0$, we have by convexity that $\mu\Phi'(\mu) - \Phi(\mu) \geq 0$, so the second term is nonnegative as well. Thus we conclude that w_r is convex in θ_r for nonzero $\boldsymbol{\theta}$, as required. □

The following lemma gives optimality conditions which characterize Nash equilibria.

Lemma 5.4 *Let $D \in \mathcal{D}(C)$ satisfy the conditions of the theorem, and let Φ be the inverse of B as given in Lemma 5.3. Let $(R, \mathbf{U})$ be a utility system. A vector $\boldsymbol{\theta} \geq 0$ is a Nash equilibrium if and only if at least two components of $\boldsymbol{\theta}$ are nonzero, and there exists a nonzero vector $\mathbf{d} \geq 0$ and a scalar $\mu > 0$ such that $\theta_r = \mu d_r$ for all r, $\sum_{r=1}^R d_r = C$, and the following conditions hold:*

$$U_r'(d_r)\left(1 - \frac{d_r}{C}\right) = \Phi(\mu)\left(1 - \frac{d_r}{C}\right) + \mu\Phi'(\mu)\left(\frac{d_r}{C}\right), \quad \textit{if } d_r > 0; \tag{5.12}$$

$$U_r'(0) \leq \Phi(\mu), \quad \textit{if } d_r = 0. \tag{5.13}$$

In this case $d_r = D(p_D(\boldsymbol{\theta}), \theta_r)$, $\mu = \sum_{r=1}^R \theta_r / C$, and $\Phi(\mu) = p_D(\boldsymbol{\theta})$.

Proof of Lemma. First suppose that $\boldsymbol{\theta}$ is a Nash equilibrium. Since $Q_r(\theta_r; \boldsymbol{\theta}_{-r}) = -\infty$ if $\boldsymbol{\theta} = 0$, (from (5.6)), we must have $\boldsymbol{\theta} \neq 0$. Suppose then that only one component of $\boldsymbol{\theta}$ is nonzero; say $\theta_r > 0$, and $\boldsymbol{\theta}_{-r} = 0$. Then the payoff to user r is:

$$U_r(C) - \Phi\left(\frac{\theta_r}{C}\right) C.$$

But now observe that by infinitesimally reducing θ_r, user r can strictly improve his payoff (since Φ is strictly increasing). Thus $\boldsymbol{\theta}$ could not have been a Nash equilibrium; we conclude that at least two components of $\boldsymbol{\theta}$ are nonzero. In this case, from (5.6), and the expressions in (5.10) and (5.11), the payoff $Q_r(\overline{\theta}_r; \boldsymbol{\theta}_{-r})$ to user r is differentiable. When two components of $\boldsymbol{\theta}$ are nonzero, we may write the payoff Q_r to user r as follows, using (5.10) and (5.11):

$$Q_r(\theta_r; \boldsymbol{\theta}_{-r}) = U_r\left(\frac{\theta_r}{\sum_{s=1}^R \theta_s} C\right) - \Phi\left(\frac{\sum_{s=1}^R \theta_s}{C}\right)\left(\frac{\theta_r}{\sum_{s=1}^R \theta_s} C\right).$$

Differentiating the previous expression with respect to θ_r, we conclude that if $\boldsymbol{\theta}$ is a Nash equilibrium then the following optimality conditions hold for each r:

$$U_r'\left(\frac{\theta_r}{\sum_{s=1}^R \theta_s}C\right)\left(\frac{C}{\sum_{s=1}^R \theta_s} - \frac{\theta_r C}{\left(\sum_{s=1}^R \theta_s\right)^2}\right) - \Phi'\left(\frac{\sum_{s=1}^R \theta_s}{C}\right)\left(\frac{\theta_r}{\sum_{s=1}^R \theta_s}\right) -$$

$$\Phi\left(\frac{\sum_{s=1}^R \theta_s}{C}\right)\left(\frac{C}{\sum_{s=1}^R \theta_s} - \frac{\theta_r C}{\left(\sum_{s=1}^R \theta_s\right)^2}\right) = 0, \quad \text{if } \theta_r > 0; \tag{5.14}$$

$$\leq 0, \quad \text{if } \theta_r = 0. \tag{5.15}$$

These conditions are equivalent to (5.12)-(5.13), if we make the substitutions $\mu = \sum_{s=1}^R \theta_s/C$, and $d_r = D(p_D(\boldsymbol{\theta}), \theta_r)$. Furthermore, in this case we have $\mathbf{d} \geq 0$, $\mu > 0$, $\theta_r = \mu d_r$, $\sum_{r=1}^R d_r = C$, and $p_D(\boldsymbol{\theta}) = \Phi(\mu)$.

On the other hand, suppose that we have found $\boldsymbol{\theta}$, $\mathbf{d}$, and μ such that the conditions of the lemma are satisfied. In this case we simply reverse the argument above; since $Q_r(\overline{\theta}_r; \boldsymbol{\theta}_{-r})$ is concave in $\overline{\theta}_r$ (Condition 2 in Definition 5.2), if at least two components of $\boldsymbol{\theta}$ are nonzero then the conditions (5.14)-(5.15) are necessary and sufficient for $\boldsymbol{\theta}$ to be a Nash equilibrium. Furthermore, if $\mathbf{d} \geq 0$, $\mu > 0$, $\theta_r = \mu d_r$, and $\sum_{r=1}^R d_r = C$, then it follows that $\mu = \sum_{s=1}^R \theta_s/C$, $\Phi(\mu) = p_D(\boldsymbol{\theta})$, and $d_r = D(p_D(\boldsymbol{\theta}), \theta_r)$. Thus the conditions (5.14)-(5.15) become equivalent to (5.12)-(5.13), as required. □

We can now use the preceding lemma to show there exists a unique Nash equilibrium, by exploiting the monotonicity of Φ; this also establishes the second claim of the theorem.

Lemma 5.5 *Let $D \in \mathcal{D}(C)$ satisfy the conditions of the theorem. Let $(R, \mathbf{U})$ be a utility system. Then there exists a unique Nash equilibrium.*

Proof of Lemma. Our approach will be to demonstrate existence of a Nash equilibrium by finding a solution $\mu > 0$ and $\mathbf{d} \geq 0$ to (5.12)-(5.13), such that $\sum_{r=1}^R d_r = C$. If we find such a solution, then at least two components of $\mathbf{d}$ must be nonzero; otherwise, (5.12) cannot hold for the user r with $d_r = C$. Thus, if we define $\boldsymbol{\theta} = \mu\mathbf{d}$, then $\mu = \sum_{s=1}^R \theta_s/C$, so $p_D(\boldsymbol{\theta}) = \Phi(\mu)$; and from (5.10), we have $d_r = D(p_D(\boldsymbol{\theta}), \theta_r)$. Thus if $\mu > 0$ and $\mathbf{d} \geq 0$ satisfy (5.12)-(5.13), then $\boldsymbol{\theta} = \mu\mathbf{d}$ is a Nash equilibrium by Lemma 5.4. Consequently, it suffices to find a solution $\mu > 0$ and $\mathbf{d} \geq 0$ to (5.12)-(5.13).

We first show that for a fixed value of $\mu > 0$, the equality in (5.12) has at most one solution d_r. To see this, rewrite (5.12) as:

$$U_r'(d_r)\left(1 - \frac{d_r}{C}\right) - (\mu\Phi'(\mu) - \Phi(\mu))\left(\frac{d_r}{C}\right) = \Phi(\mu).$$

Since Φ is convex and strictly increasing with $\Phi(\mu) \to 0$ as $\mu \to 0$, we have $\mu\Phi'(\mu) - \Phi(\mu) \geq 0$. Thus the left hand side is strictly decreasing in d_r (since U_r is strictly increasing and concave), from $U_r'(0)$ at $d_r = 0$ to $\Phi(\mu) - \mu\Phi'(\mu) \leq 0$ when $d_r = C$. This implies a unique solution $d_r \in [0, C]$ exists for the equality in (5.12) as long as $U_r'(0) \geq \Phi(\mu)$; we denote this solution $d_r(\mu)$. If $\Phi(\mu) > U_r'(0)$, then we let $d_r(\mu) = 0$. Observe that as $\mu \to 0$, we must have $d_r(\mu) \to C$, since otherwise we can show that (5.12) fails to hold for sufficiently small μ.

Next we show that $d_r(\mu)$ is continuous. Since we defined $d_r(\mu) = 0$ if $\Phi(\mu) > U_r'(0)$, and $d_r(\mu) = 0$ if $\Phi(\mu) = U_r'(0)$ from (5.12), it suffices to show that $d_r(\mu)$ is continuous for μ such that $\Phi(\mu) \leq U_r'(0)$. But in this case continuity of d_r can be shown using (5.12), together with the fact that U_r', Φ, and Φ' are all continuous (the latter because Φ is concave and differentiable, and hence continuously differentiable). Indeed, suppose that $\mu_n \to \mu$ where $\Phi(\mu) \leq U_r'(0)$, and assume without loss of generality that $d_r(\mu_n) \to d_r$ (since $d_r(\mu_n)$ takes values in the compact set $[0, C]$). Then since μ_n and $d_r(\mu_n)$ satisfy the equality in (5.12) for sufficiently large n, by taking limits we see that μ and d_r satisfy the equality in (5.12) as well. Thus we must have $d_r = d_r(\mu)$, so we conclude $d_r(\mu)$ is continuous.

We now show that $d_r(\mu)$ is nonincreasing in μ. To see this, choose $\mu_1, \mu_2 > 0$ such that $\mu_1 < \mu_2$. Suppose that $d_r(\mu_1) < d_r(\mu_2)$. Then, in particular, $d_r(\mu_2) > 0$, so (5.12) holds with equality for $d_r(\mu_2)$ and μ_2. Now note that as we move from $d_r(\mu_2)$ to $d_r(\mu_1)$, the left hand side of (5.12) strictly increases (since U_r is concave). On the other hand, since Φ is convex and strictly increasing with $\Phi(\mu) \to 0$ as $\mu \to 0$, we have the inequalities $\mu_2\Phi'(\mu_2) - \Phi(\mu_2) \geq \mu_1\Phi'(\mu_1) - \Phi(\mu_1) \geq 0$. From this it follows that the right hand side of (5.12) strictly decreases as we move from $d_r(\mu_2)$ to $d_r(\mu_1)$ and from μ_2 to μ_1. Thus neither (5.12) nor (5.13) can hold at $d_r(\mu_1)$ and μ_1; so we conclude that for all r, we must have $d_r(\mu_1) \geq d_r(\mu_2)$.

Thus for each r, $d_r(\mu)$ is a nonincreasing continuous function such that $d_r(\mu) \to C$ as $\mu \to 0$, and $d_r(\mu) \to 0$ as $\mu \to \infty$. We conclude there exists at least one $\mu > 0$ such that $\sum_{r=1}^{R} d_r(\mu) = C$; and in this case $\mathbf{d}(\mu)$ satisfies (5.12)-(5.13), so by the discussion at the beginning of this proof, we know that $\boldsymbol{\theta} = \mu\mathbf{d}(\mu)$ is a Nash equilibrium.

Finally, we show that the Nash equilibrium is unique. Suppose that there exist two solutions $\mathbf{d}^1 \geq 0, \mu_1 > 0$, and $\mathbf{d}^2 \geq 0, \mu_2 > 0$ to (5.12)-(5.13), such that $\sum_{r=1}^{R} d_r^i = C$ for $i = 1, 2$. Of course, we must have $\mathbf{d}^i = \mathbf{d}(\mu_i)$, $i = 1, 2$. We assume without loss of generality that $\mu_1 \leq \mu_2$; our goal is to show that $\mu_1 = \mu_2$. Since $d_r(\cdot)$ is nonincreasing, we know $d_r(\mu_1) \geq d_r(\mu_2)$ for all r. Suppose that $d_r(\mu_1) > d_r(\mu_2)$ for all r with $d_r(\mu_2) > 0$. Then we have:

$$C = \sum_{r=1}^{R} d_r^1 = \sum_{r=1}^{R} d_r(\mu_1) > \sum_{r=1}^{R} d_r(\mu_2) = \sum_{r=1}^{R} d_r^2 = C,$$

a contradiction. Thus we must have $d_r(\mu_1) = d_r(\mu_2)$ for some user r with $d_r(\mu_2) > 0$. Observe that $\Phi(\mu)$ and $\mu\Phi'(\mu)$ are both strictly increasing in $\mu > 0$, since Φ is strictly increasing and convex. Thus for fixed $d_r > 0$, the equality in (5.12) has a unique solution μ, so $d_r(\mu_1) = d_r(\mu_2) > 0$ implies $\mu_1 = \mu_2$, in which case $\mathbf{d}^1 = \mathbf{d}^2$. Thus (5.12)-(5.13) have a unique solution $\mathbf{d} \geq 0$, $\mu > 0$, such that $\sum_{r=1}^{R} d_r = C$. From Lemma 5.4, this ensures the Nash equilibrium $\boldsymbol{\theta} = \mu\mathbf{d}$ is unique as well. □

We now proceed to compute the efficiency loss of mechanisms in $\mathcal{D}(C)$ satisfying the conditions of the theorem. Following the same reasoning as in Lemma 2.7, it is straightforward to show that the worst case efficiency loss occurs when all utility functions are linear. We assume without loss of generality that $U_r(d_r) = \alpha_r d_r$, for $r = 1, \ldots, R$. Let $A = \max_r \alpha_r$; relabeling if necessary, we also assume that $\alpha_1 = A$. Then the optimal solution to *SYSTEM*$(C, R, \mathbf{U})$ has optimal value AC. We now proceed to compute the worst case aggregate utility at a Nash equilibrium.

To identify the worst case situation, we would like to find $\alpha_2, \ldots, \alpha_R \leq A$ such that $Ad_1 + \sum_{r=2}^{R} \alpha_r d_r$, the total utility associated with the Nash equilibrium, is as small as possible; this results in the following optimization problem (with unknowns $\mu, d_1, \ldots, d_R, \alpha_2, \ldots, \alpha_R$):

$$\text{minimize} \quad Ad_1 + \sum_{r=2}^{R} \alpha_r d_r \tag{5.16}$$

$$\text{subject to} \quad \alpha_r \left(1 - \frac{d_r}{C}\right) = \Phi(\mu)\left(1 - \frac{d_r}{C}\right) + \mu\Phi'(\mu)\left(\frac{d_r}{C}\right), \qquad \text{if } d_r > 0; \tag{5.17}$$

$$\alpha_r \leq \Phi(\mu), \qquad \text{if } d_r = 0; \tag{5.18}$$

$$\sum_r d_r = C; \tag{5.19}$$

$$0 < \alpha_r \leq A, \qquad r = 2, \ldots, R; \tag{5.20}$$

$$\mu > 0; \; d_r \geq 0, \qquad r = 1, \ldots, R. \tag{5.21}$$

As in the proof of Theorem 2.6, this optimization problem *chooses* linear utility functions with slopes less than or equal to A for players $2, \ldots, R$. The constraints in the problem require that given linear utility functions $U_r(d_r) = \alpha_r d_r$ for $r = 1, \ldots, R$, the allocation $\mathbf{d}$ must in fact be the unique Nash equilibrium allocation of the resulting game. (The resulting Nash equilibrium $\boldsymbol{\theta}$ is given by $\boldsymbol{\theta} = \mu\mathbf{d}$, as in Lemma 5.4.) As a result, the optimal objective function value is precisely the lowest possible aggregate utility achieved, among all such games; since the optimal value of *SYSTEM*$(C, R, \mathbf{U})$ is fixed at AC, this computation will yield $\rho(C, D)$ as well.

Suppose now $(\mu, \mathbf{d}, \boldsymbol{\alpha})$ is an optimal solution to (5.16)-(5.21). If $\Phi(\mu) \geq A$, it follows from (5.17)-(5.18) that $d_r = 0$ for all r (since $\mu\Phi'(\mu) - \Phi(\mu) \geq 0$); thus (5.19) cannot hold.

We conclude that $\Phi(\mu) < A$, in which case (5.17)-(5.18) imply that $d_1 > 0$, since $\alpha_1 = A$. Now suppose that $n < R$ users, say users $r = R - n + 1, \dots, R$, have $d_r = 0$. Then the first $R - n$ coordinates of $\boldsymbol{\alpha}$ and $\mathbf{d}$ must be an optimal solution to the problem (5.16)-(5.21), with $R-n$ users. Therefore, in finding the worst case game, it suffices to assume that $d_r > 0$ for all $r = 1, \dots, R$, and then consider the optimal objective function value for $R = 2, 3, \dots$. This allows us to replace the pair of constraints (5.17)-(5.18) with the following single constraint:

$$\alpha_r \left(1 - \frac{d_r}{C}\right) = \Phi(\mu)\left(1 - \frac{d_r}{C}\right) + \mu\Phi'(\mu)\left(\frac{d_r}{C}\right). \tag{5.22}$$

We now fix $\mu > 0$. Since $\alpha_1 = A$, (5.22) implies that:

$$d_1 = \frac{(A - \Phi(\mu))C}{A - \Phi(\mu) + \mu\Phi'(\mu)}. \tag{5.23}$$

The constraint $d_1 > 0$ becomes equivalent to $\Phi(\mu) < A$, since $\mu\Phi'(\mu) - \Phi(\mu) \geq 0$.

Furthermore, given $\mu > 0$, we can rewrite (5.22) as:

$$\alpha_r = \frac{\Phi(\mu)C + (\mu\Phi'(\mu) - \Phi(\mu))d_r}{C - d_r}. \tag{5.24}$$

The constraint $\alpha_r \leq A$ can now be rewritten as:

$$d_r \leq \frac{(A - \Phi(\mu))C}{A - \Phi(\mu) + \mu\Phi'(\mu)}. \tag{5.25}$$

(Note the right hand side is exactly equal to d_1, from (5.23).)

Thus we have the following "reduced" optimization problem:

$$\text{minimize} \quad \frac{(A - \Phi(\mu))AC}{A - \Phi(\mu) + \mu\Phi'(\mu)} + \sum_{r=2}^{R} \frac{\Phi(\mu)Cd_r + (\mu\Phi'(\mu) - \Phi(\mu))d_r^2}{C - d_r} \tag{5.26}$$

$$\text{subject to} \quad \sum_{r=2}^{R} d_r = C - \frac{(A - \Phi(\mu))C}{A - \Phi(\mu) + \mu\Phi'(\mu)}; \tag{5.27}$$

$$d_r \leq \frac{(A - \Phi(\mu))C}{A - \Phi(\mu) + \mu\Phi'(\mu)}, \quad r = 2, \dots, R; \tag{5.28}$$

$$\Phi(\mu) < A; \tag{5.29}$$

$$\mu > 0; \; d_r > 0, \quad r = 2, \dots, R. \tag{5.30}$$

The objective function (5.26) is equivalent to (5.16), upon substituting with (5.23) and (5.24); these substitutions also ensure that (5.17)-(5.18) hold. The constraint (5.27) is equivalent to (5.19), upon substitution from (5.23). The constraint $\alpha_r > 0$ in (5.20) holds from (5.24), since (5.27) and (5.30) ensure that $0 < d_r < C$. The constraint $\alpha_r \le 1$ in (5.20) becomes equivalent to (5.28) (see (5.25)). Finally, the constraint that $d_1 > 0$ yields (5.29), from (5.23).

For fixed $\mu > 0$, it follows that there exists a feasible solution to (5.26)-(5.30) if and only if $\Phi(\mu) < A$, and R is sufficiently large, i.e.:

$$\frac{C}{R} \le \frac{(A - \Phi(\mu))C}{A - \Phi(\mu) + \mu\Phi'(\mu)}. \tag{5.31}$$

In this case, the following symmetric solution is feasible:

$$d_r = \left(\frac{1}{R-1}\right)\left(C - \frac{(A - \Phi(\mu))C}{A - \Phi(\mu) + \mu\Phi'(\mu)}\right), \quad r = 2, \ldots, R. \tag{5.32}$$

Notice that the objective function (5.26) is strictly convex and symmetric in d_r; thus the feasible solution (5.32) must in fact be an optimal solution to (5.26)-(5.30). If we substitute this solution into the objective function (5.26), the resulting expression is decreasing in R. Thus the worst case efficiency loss occurs as $R \to \infty$. Furthermore, for fixed $\mu > 0$ such that $\Phi(\mu) < A$, as $R \to \infty$ the constraint (5.31) holds for all sufficiently large R. We conclude the worst case aggregate utility at a Nash equilibrium is given by solving the following optimization problem:

$$\text{minimize} \quad \frac{(A - \Phi(\mu))AC}{A - \Phi(\mu) + \mu\Phi'(\mu)} + \left(C - \frac{(A - \Phi(\mu))C}{A - \Phi(\mu) + \mu\Phi'(\mu)}\right)\Phi(\mu) \tag{5.33}$$

$$\text{subject to} \quad \Phi(\mu) < A; \ \mu > 0. \tag{5.34}$$

The first term of (5.33) is identical to the first term of (5.26). The second term of (5.33) is the limit as $R \to \infty$ of the expression $\sum_{r=2}^{R} \alpha_r d_r$, where α_r is defined in (5.24) and d_r is defined in (5.32). The limit follows since $d_r \to 0$ as $R \to \infty$, so $\alpha_r \to \Phi(\mu)$.

The optimization problem (5.33)-(5.34) gives the worst case aggregate utility at a Nash equilibrium. Recall that since all users have linear utility functions with slopes less than or equal to A, while user 1 has marginal utility exactly equal to A, the maximal aggregate utility is equal to AC. Thus the worst case *efficiency loss* at a Nash equilibrium relative to the maximal aggregate utility is given by solving the following problem, with unknowns $\mu > 0, A > 0$:

$$\text{minimize} \quad \frac{(A - \Phi(\mu))}{A - \Phi(\mu) + \mu\Phi'(\mu)} + \left(1 - \frac{(A - \Phi(\mu))}{A - \Phi(\mu) + \mu\Phi'(\mu)}\right)\left(\frac{\Phi(\mu)}{A}\right) \tag{5.35}$$

$$\text{subject to} \quad \Phi(\mu) < A; \ \mu > 0. \tag{5.36}$$

The objective function (5.35) results by dividing the objective function (5.33) through by AC.

We briefly summarize the sequence of reductions made to this point. Suppose that $(R, \mathbf{U})$ is a utility system, and let $\boldsymbol{\theta}$ be a Nash equilibrium. Define A as the largest marginal utility at the Nash equilibrium, $A = \max_r U_r'(D(p_D(\boldsymbol{\theta}), \theta_r)) > 0$. In addition, let $\mu = \sum_{s=1}^R \theta_s / C$. Then the ratio of Nash equilibrium aggregate utility to the maximal aggregate utility is no worse than the value of the objective function (5.35) given A and μ. Furthermore, for fixed $\mu > 0$ and A such that $A > \Phi(\mu)$, there exists a sequence of games (where $R \to \infty$) such that the ratio of Nash equilibrium aggregate utility to the maximal aggregate utility approaches the value of the objective function (5.35). Thus, if we solve the optimization problem (5.35)-(5.36), the resulting optimal value is exactly equal to $\rho(C, D)$.

If we rearrange the terms of (5.35) and make the substitutions $x = \Phi(\mu)/A$, and $\Psi(\mu) = \mu\Phi'(\mu)/\Phi(\mu) \geq 1$, we have the following equivalent optimization problem:

$$\text{minimize} \quad \frac{(1-x)^2}{1 + (\Psi(\mu) - 1)x} + x \tag{5.37}$$

$$\text{subject to} \quad 0 < x < 1;\ \mu > 0. \tag{5.38}$$

For fixed $\mu > 0$, denote the objective function value (5.37) by $F(x; \mu)$:

$$F(x; \mu) = \frac{(1-x)^2}{1 + (\Psi(\mu) - 1)x} + x. \tag{5.39}$$

Observe that for fixed $\mu > 0$, $F(x; \mu)$ is strictly convex, and approaches 1 either as $x \to 0$ or as $x \to 1$. Thus for fixed $\mu > 0$ there exists a unique optimal solution $x \in (0, 1)$ to (5.37)-(5.38) in $(0, 1)$, which we denote $x^*(\mu)$. This optimal solution is straightforward to identify by differentiating $F(x; \mu)$ with respect to x; we have:

$$x^*(\mu) = \begin{cases} \dfrac{1}{2}, & \text{if } \Psi(\mu) = 1; \\[2ex] \dfrac{\sqrt{\Psi(\mu)} - 1}{\Psi(\mu) - 1}, & \text{if } \Psi(\mu) > 1. \end{cases}$$

After substituting and simplifying, it is straightforward to verify that $F(x^*(\mu); \mu) = G(\Psi(\mu))$, where $G(\Psi)$ is defined for $\Psi \geq 1$ according to:

$$G(\Psi) = \begin{cases} \dfrac{3}{4}, & \text{if } \Psi = 1; \\[2ex] \dfrac{2\Psi^2 - 3\Psi\sqrt{\Psi} + \sqrt{\Psi}}{(\Psi - 1)^2\sqrt{\Psi}}, & \text{if } \Psi > 1. \end{cases} \tag{5.40}$$

Since $\rho(C, D)$ is the optimal value of the optimization problem (5.35)-(5.36), we conclude:

$$\rho(C, D) = \inf_{\mu > 0} F(x^*(\mu), \mu) = \inf_{\mu > 0} G(\Psi(\mu)).$$

From the definition, it is straightforward to verify that $G(\Psi)$ is continuous, by checking that $G(\Psi) \to 3/4$ as $\Psi \to 1$. If $\Psi(\mu_1) > \Psi(\mu_2) \geq 1$, then $F(x; \mu_1) < F(x; \mu_2)$ for all x such that $0 < x < 1$; this follows directly from the definition of F in (5.39). Thus we must have $G(\Psi(\mu_1)) < G(\Psi(\mu_2))$. We conclude $G(\Psi)$ is strictly decreasing, and that $G(\Psi)$ approaches zero as $\Psi \to \infty$ (using the definition (5.40)). Thus if $\sup_{\mu > 0} \Psi(\mu) = \infty$, then $\rho(C, D) = \inf_{\mu > 0} G(\Psi(\mu)) = 0$. On the other hand, if $\sup_{\mu > 0} \Psi(\mu) = \Psi^* < \infty$, then $\rho(C, D) = \inf_{\mu > 0} G(\Psi(\mu)) = G(\Psi^*)$. Note that since B is the differentiable inverse of Φ on $(0, \infty)$, we have:

$$\sup_{\mu > 0} \Psi(\mu) = \sup_{\mu > 0} \left[\frac{\mu \Phi'(\mu)}{\Phi(\mu)} \right] = \left(\inf_{p > 0} \left[\frac{p B'(p)}{B(p)} \right] \right)^{-1},$$

where we interpret the right hand side as infinity if the infimum is equal to zero. Furthermore, for $0 < \varepsilon \leq 1$, we have $g(\varepsilon) = G(1/\varepsilon)$, where g is defined as in (5.7). Thus we have $\rho(C, D) = g(\varepsilon^*)$, where ε^* is defined in (5.8).

Finally, suppose that $\Phi(\mu) = \mu \Phi'(\mu)$ for all $\mu > 0$. Then $\Psi(\mu) = 1$ for all $\mu > 0$, so $G(\Psi(\mu)) = 3/4$ for all $\mu > 0$, and $\rho(C, D) = 3/4$. Since Φ is convex and strictly increasing, with $\Phi(\mu) \to 0$ as $\mu \to 0$, we can only have $\Psi(\mu) = 1$ for all $\mu > 0$ if $\Phi(\mu) = \Delta\mu$ for some $\Delta > 0$; thus $B(p) = p/\Delta$, and we conclude that $D(p, \theta) = \Delta\theta/p$, as required. On the other hand, if $\mu \Phi'(\mu) > \Phi(\mu)$ for some $\mu > 0$, then $\Psi(\mu) > 1$, and $G(\Psi(\mu)) < 3/4$, so $\rho(C, D) < 3/4$. This establishes the third claim of the theorem. □

The preceding theorem establishes that the mechanism proposed in Section 2.1 minimizes efficiency loss among a wide class of mechanisms with certain desirable properties. In fact, the theorem proves something much stronger: we explicitly show that all mechanisms D satisfying the conditions of the theorem must be of the form $D(p, \theta) = \theta/B(p)$. Furthermore, we show that the worst case efficiency loss of such a mechanism is governed by the degree of "nonlinearity" of $B(p)$, as measured through the quantity ε^* defined in (5.8). Note that the quantity $pB'(p)/B(p)$ is the *elasticity* of $B(p)$ [137]; thus ε^* is the minimal elasticity of $B(p)$ over all $p > 0$.

We note that one potentially undesirable feature of the family of market-clearing mechanisms considered is that the payoff to user r is defined as $-\infty$ when the composite strategy vector is $\boldsymbol{\theta} = 0$ (cf. (5.6)). This definition is required because when the composite strategy vector is $\boldsymbol{\theta} = 0$, a market-clearing price may not exist. One possible remedy is to restrict attention instead to mechanisms where $D(p, \theta) = 0$ if $\theta = 0$, for all $p \geq 0$; in this case we can *define* $p_D(\boldsymbol{\theta}) = 0$ if $\boldsymbol{\theta} = 0$, and let the payoff to user r be $U_r(0)$ if $\theta_r = 0$. This condition amounts to a "normalization" on the market-clearing mechanism. Furthermore, this modification now captures the mechanism of Chapter 2, where $Q_r(0; \mathbf{w}_{-r}) = U_r(0)$ for all $\mathbf{w}_{-r} \geq 0$ (see (2.13)). It is straightforward to check that this modification does not alter the conclusion of Theorem 5.1, since the class of mechanisms in $\mathcal{D}(C)$ satisfying the conditions of Theorem 5.1 can be extended to ensure $D(p, \theta) = 0$ if $\theta = 0$.

Finally, we note one desirable feature of the mechanisms considering in Theorem 5.1, particularly in the context of communication networks. In general, even though the strategy space of the users is one-dimensional, the market-clearing price $p_D(\boldsymbol{\theta})$ may have a complex dependence on the vector $\boldsymbol{\theta}$. However, from Lemma 5.3, note that the market clearing price $p_D(\boldsymbol{\theta})$ depends only on the *sum* of the strategies of all the players, $\sum_s \theta_s$. Thus, under the conditions of Theorem 5.1, we show that the market-clearing price is only a function of a simple aggregate $\sum_s \theta_s$ of the players' strategies, so that the market-clearing process does not require identification of individual players interacting with the mechanism, or even the number of players. This is a desirable scaling property for market mechanisms to be deployed in large scale networks.

■ 5.1.2 A Second Characterization Theorem

Note that given the definition of $\mathcal{D}(C)$, we have considered market-clearing mechanisms for a fixed supply C. Intuitively, if we require a mechanism to be well defined for all capacities C, then given Condition 2 in the definition of $\mathcal{D}(C)$ one might expect Condition 1 to hold as well; that is, for an appropriate limiting case we expect that price anticipating users will be approximately price taking.

Formally, we define the class $\hat{\mathcal{D}}$ as follows.

Definition 5.3
The class $\hat{\mathcal{D}}$ consists of all functions $D(p, \theta)$ such that the following conditions are satisfied:

1. *For all $C > 0$, D is a smooth market-clearing mechanism for C (cf. Definition 5.1).*

2. *For all $C > 0$, and for all $U_r \in \mathcal{U}$, a user's payoff is concave if he is price anticipating; that is, for all R, and for all $\boldsymbol{\theta}_{-r} \in (\mathbb{R}^+)^R$, the function:*

$$U_r(D(p_D(\boldsymbol{\theta}), \theta_r) - p_D(\boldsymbol{\theta}) D(p_D(\boldsymbol{\theta}), \theta_r)$$

   is concave in $\theta_r > 0$ if $\boldsymbol{\theta}_{-r} = \mathbf{0}$, and concave in $\theta_r \geq 0$ if $\boldsymbol{\theta}_{-r} \neq \mathbf{0}$.

3. *The demand functions are nonnegative; i.e., for all $p > 0$ and $\theta \geq 0$, $D(p, \theta) \geq 0$.*

Note that any mechanism in $\hat{\mathcal{D}}$ must be a smooth market-clearing mechanism for any $C > 0$; in particular, the market-clearing price $p_D(\boldsymbol{\theta})$ must be uniquely defined for any $C > 0$. (Note that in the notation we suppress the dependence of the market-clearing price $p_D(\boldsymbol{\theta})$ on the capacity C.) We have the following proposition.

Proposition 5.6
If $D \in \hat{\mathcal{D}}$, then for all $C > 0$, $D \in \mathcal{D}(C)$.

Proof. Fix $D \in \hat{\mathcal{D}}$. It suffices to show that Condition 1 in Definition 5.2 is satisfied. Our approach is to show that for an appropriate limiting environment, price anticipating users behave as if they are price taking.

First suppose that for fixed $\theta > 0$, there exist $\mu_1, \mu_2 > 0$ with $\mu_1 \neq \mu_2$ such that $D(\mu_1, \theta) = D(\mu_2, \theta) = d$. If $d > 0$, then let $C = 2d$ and let $R = 2$. Then for $\boldsymbol{\theta} = (\theta, \theta)$, there cannot exist a unique market-clearing price $p_D(\boldsymbol{\theta})$. We conclude that $D(\cdot, \theta)$ must be a strictly monotonic function for $\mu > 0$; furthermore, the condition that a unique market-clearing price must exist implies that as $\mu \to 0$ or $\mu \to \infty$, either $D(\mu, \theta) \to 0$ or $D(\mu, \theta) \to \infty$.

Let $I \subset (0, \infty)$ be the set of $\theta > 0$ such that $D(\mu, \theta)$ is strictly increasing in μ; note that if $\theta \in (0, \infty) \setminus I$, then $D(\mu, \theta)$ is necessarily strictly decreasing in μ. Suppose $I \neq (0, \infty)$ and $I \neq \emptyset$; then choose $\theta \in \partial I$, the boundary of I. Choose a sequence $\theta_n \in I$ such that $\theta_n \to \theta$; and choose another sequence $\hat{\theta}_n \in (0, \infty) \setminus I$ such that $\hat{\theta}_n \to \theta$. Fix μ_1, μ_2 with $0 < \mu_1 < \mu_2$. Then we have $D(\mu_1, \theta_n) < D(\mu_2, \theta_n)$, and $D(\mu_1, \hat{\theta}_n) > D(\mu_2, \hat{\theta}_n)$. Taking limits as $n \to \infty$, we get $D(\mu_1, \theta) \leq D(\mu_2, \theta)$, and $D(\mu_1, \theta) \geq D(\mu_2, \theta)$, so that $D(\mu_1, \theta) = D(\mu_2, \theta)$. But this is not possible, as shown above (since $D(\cdot, \theta)$ must be strictly monotonic). Thus $I = (0, \infty)$ or $I = \emptyset$. In particular, either $D(\mu, \theta)$ is strictly decreasing in $\mu > 0$ for all $\theta > 0$, or strictly increasing in $\mu > 0$ for all $\theta > 0$.

As in Lemma 5.2, it is straightforward to check that for all C, we must have that $D(p_D(\boldsymbol{\theta}), \theta_r)$ is concave in θ_r, for nonzero $\boldsymbol{\theta}$. We will use this fact to show $D(\mu, \theta)$ is concave in $\theta \geq 0$ for fixed $\mu > 0$. Since $D(\mu, \theta)$ is continuous, it suffices to show that $D(\mu, \theta)$ is concave for $\theta > 0$. Suppose not; fix $\theta > 0$, $\overline{\theta} > 0$, and $\delta \in (0, 1)$ such that:

$$D(\mu, \delta\theta + (1-\delta)\overline{\theta}) < \delta D(\mu, \theta) + (1-\delta) D(\mu, \overline{\theta}). \tag{5.41}$$

Note this implies in particular that either $D(\mu, \theta) > 0$ or $D(\mu, \overline{\theta}) > 0$. We assume without loss of generality that $D(\mu, \theta) > 0$. Let $C^R = RD(\mu, \theta)$, and let $\boldsymbol{\theta}^R = (\theta, \ldots, \theta) \in (\mathbb{R}^+)^R$. To emphasize the dependence of the market-clearing price on the capacity, we will let $p_D(\overline{\boldsymbol{\theta}}; C)$ denote the market-clearing price when the composite strategy vector is $\overline{\boldsymbol{\theta}}$ and the capacity is C. We will show that for any $\theta' > 0$, if $\mu^R = p_D(\boldsymbol{\theta}^{R-1}, \theta'; C^R)$, then $\mu^R \to \mu$ as $R \to \infty$. First note that by definition, we have $D(\mu^R, \theta') + (R-1)D(\mu^R, \theta) = RD(\mu, \theta)$; or, rewriting, we have:

$$\frac{1}{R} D(\mu^R, \theta') + \left(1 - \frac{1}{R}\right) D(\mu^R, \theta) = D(\mu, \theta). \tag{5.42}$$

Now note that as $R \to \infty$, the right hand side remains constant. Suppose that $\mu^R \to \infty$. Since $I = (0, \infty)$ or $I = \emptyset$, either $D(\mu^R, \theta'), D(\mu^R, \theta) \to 0$, or $D(\mu^R, \theta'), D(\mu^R, \theta) \to \infty$;

in either case, the equality (5.42) is violated for large R. A similar conclusion holds if $\mu^R \to 0$ as $R \to \infty$. Thus we do not have $\mu^R \to 0$ or $\mu^R \to \infty$ as $R \to \infty$. Choose a convergent subsequence R_k, such that $\mu^{R_k} \to \hat{\mu}$, where $\hat{\mu} \in (0, \infty)$. From (5.42), we must have $D(\hat{\mu}, \theta) = D(\mu, \theta)$. But as established above, since $D(\cdot, \theta)$ is strictly monotonic, this is only possible if $\hat{\mu} = \mu$. We conclude that the following three limits hold:

$$\lim_{R\to\infty} p_D(\boldsymbol{\theta}^R; C^R) = \mu;$$
$$\lim_{R\to\infty} p_D(\boldsymbol{\theta}^{R-1}, \overline{\theta}; C^R) = \mu;$$
$$\lim_{R\to\infty} p_D(\boldsymbol{\theta}^{R-1}, \delta\theta + (1-\delta)\overline{\theta}; C^R) = \mu.$$

The remainder of the proof is straightforward. From (5.41), for R sufficiently large, we must have:

$$D(p_D(\boldsymbol{\theta}^{R-1}, \delta\theta + (1-\delta)\overline{\theta}; C^R), \delta\theta + (1-\delta)\overline{\theta}) < \delta D(p_D(\boldsymbol{\theta}^R; C^R), \theta) + (1-\delta) D(p_D(\boldsymbol{\theta}^{R-1}, \overline{\theta}; C^R), \overline{\theta}).$$

This violates Condition 1 in the definition of $\hat{\mathcal{D}}$. A similar argument shows that $\mu D(\mu, \theta)$ is convex in θ, by using the fact that $p_D(\boldsymbol{\theta}) D(p_D(\boldsymbol{\theta}), \theta_r)$ must be convex in θ_r for nonzero $\boldsymbol{\theta}$ (which follows using the same logic as Lemma 5.2). We conclude that Condition 1 of the definition of $\mathcal{D}(C)$ holds, so $D \in \mathcal{D}(C)$, as required. □

From the preceding proposition and Theorem 5.1, it follows that for $D \in \hat{\mathcal{D}}$, the same result as Theorem 5.1 holds.

Theorem 5.7

Let $D \in \hat{\mathcal{D}}$ be a smooth market-clearing mechanism such that for all capacities $C > 0$ and utility systems $(R, \mathbf{U})$, there exists a competitive equilibrium $\boldsymbol{\theta}$ such that $(D(p_D(\boldsymbol{\theta}), \theta_r), r = 1, \ldots, R)$ solves SYSTEM$(C, R, \mathbf{U})$. Then the conclusions of Theorem 5.1 hold for D; in particular, for any capacity C and utility system $(R, \mathbf{U})$, there exists a unique Nash equilibrium. Furthermore, $\rho(C, D) \leq 3/4$, and this bound is met with equality if and only if $D(p, \theta) = \Delta\theta/p$ for some $\Delta > 0$.

We comment briefly here on the possibility of relaxing Condition 2 in the definitions of $\mathcal{D}(C)$ and $\hat{D}$: that users' payoffs are concave when they are price anticipating. We motivated this assumption in Section 5.1.1 as a means of characterization of pure strategy Nash equilibria. However, if we relax this assumption, then in general one can still expect existence of *mixed* strategy Nash equilibria [96], where the strategy of a player is a probability distribution over available actions. In this case one can then ask whether the *expected* aggregate utility of a mixed strategy Nash equilibrium suffers a high efficiency loss relative to the maximal aggregate utility. However, such a direction constitutes a significant departure from the material of this section as well as the previous section, because we use Condition 2 in the definitions of $\mathcal{D}(C)$ and $\hat{D}$ not only to characterize pure strategy Nash equilibria, but also to constrain the structure of the mechanisms under consideration.

The theorems of this section and the previous section suggest that the best efficiency guarantee we can hope to achieve is 75%, if we are restricted to market-clearing mechanisms with scalar strategy spaces. Mechanisms which do not choose a single price to clear the market can lead to lower efficiency losses. For example, Sanghavi and Hajek [112] have shown that if users choose their total payments, but the link manager is allowed to choose the allocation to users as an arbitrary function of the payments, it is possible to ensure no worse than a 13% efficiency loss. Furthermore, Yang and Hajek [148] have shown that if a mechanism allocates resources in proportion to the users' strategies (i.e., user r receives a fraction $\theta_r/(\sum_{s=1}^R \theta_s)$ of the resource), then by using differentiated pricing, it is possible to guarantee arbitrarily small efficiency loss at the Nash equilibrium. Note that, in light of Lemma 5.3, any mechanism in $\mathcal{D}(C)$ satisfying the conditions of Theorem 5.1 must in fact allocate resources in proportion to the users' strategies as well. However, we can only guarantee an efficiency loss of no more than 25%, because the mechanisms in $\mathcal{D}(C)$ set only a single price, and choose this price as a function of $\sum_{r=1}^R \theta_r$ (as shown in Lemma 5.3). Indeed, in the next section, we will present a two-user mechanism which can guarantee an arbitrarily low efficiency loss by setting a single price, because the conditions imposed in the analysis of this section are violated.

5.1.3 A Two User Mechanism with Arbitrarily Low Efficiency Loss

In this section we consider the special case where $R = 2$. We develop a mechanism where the strategy space of each user is $\mathbb{R}^+$, and where the mechanism sets a single price, such that the efficiency loss when users are price anticipating can be guaranteed to be arbitrarily low. However, we will see that such a mechanism *cannot* be defined in terms of users submitting demand functions, with a price being chosen to "clear the market."

The mechanism is defined as follows. Let the capacity of the resource be $C > 0$, and assume the two users have utility functions $U_1, U_2 \in \mathcal{U}$. Fix γ such that $0 < \gamma < 1$. Each user chooses a strategy $\theta_i \geq 0$, $i = 1, 2$; the payoff to user 1 is then:

$$Q_1(\theta_1;\theta_2) = \begin{cases} U_1\left(\dfrac{(\theta_1 - \gamma\theta_2)^+}{(\theta_1 - \gamma\theta_2)^+ + (\theta_2 - \gamma\theta_1)^+} C\right) - (\theta_1 - \gamma\theta_2)^+, & \text{if } \theta_1 > 0; \\ \\ U_1(0), & \text{if } \theta_1 = 0. \end{cases} \tag{5.43}$$

The payoff to user 2 is defined symmetrically. We interpret the preceding mechanism as follows. Given the strategy vector $\boldsymbol{\theta}$, the resource manager chooses a price $p(\boldsymbol{\theta})$ given by:

$$p(\boldsymbol{\theta}) = \frac{(\theta_1 - \gamma\theta_2)^+ + (\theta_2 - \gamma\theta_1)^+}{C}.$$

If $p(\boldsymbol{\theta}) > 0$, then user 1 is allocated a rate given by $d_1(\boldsymbol{\theta}) = w_1(\boldsymbol{\theta})/p(\boldsymbol{\theta})$, where $w_1(\boldsymbol{\theta}) = (\theta_1 - \gamma\theta_2)^+$ is the payment made by user 1. The same definitions hold symmetrically for user 2. We note that in the special case where $\gamma = 0$, this mechanism is identical to the mechanism studied in Section 2.1.

Note that in this case, the resource manager is choosing a single price $p(\boldsymbol{\theta})$ which is charged to both users. However, we cannot interpret this game as one where users submit demand functions; to see this, note that the "demand" of user 1 at a given price p is given by $(\theta_1 - \gamma\theta_2)^+/p$. However, this is not a function only of θ_1 and p; it also depends on the strategy of the second player, θ_2. For this reason the mechanism does not fit the definition of market-clearing mechanism outlined in Section 5.1.1. In fact, the mechanism also exhibits a more striking property: in general, the payoffs Q_r are not concave in the strategy θ_r. Nevertheless, the following theorem shows that a Nash equilibrium always exists, and as $\gamma \to 1$, the efficiency loss approaches zero.

Theorem 5.8

Fix $U_1, U_2 \in \mathcal{U}$ and $C > 0$. For any γ such that $0 < \gamma < 1$, there exists a Nash equilibrium (θ_1, θ_2) of the game defined by (Q_1, Q_2). Furthermore, all such Nash equilibria lead to the same allocation to the two players. If we denote this allocation by $(d_1(\gamma), d_2(\gamma))$, and let (d_1^S, d_2^S) be any optimal solution to SYSTEM$(C, 2, \mathbf{U})$, then there holds:

$$\lim_{\gamma \to 1} \frac{U_1(d_1(\gamma)) + U_2(d_2(\gamma))}{U_1(d_1^S) + U_2(d_2^S)} = 1.$$

Proof. Fix $\theta_2 \geq 0$. We start by rewriting the payoff to user 1 as follows:

$$Q_1(\theta_1; \theta_2) = \begin{cases} U_1(0), & \text{if } 0 \leq \theta_1 < \gamma\theta_2; \\ \\ U_1\left(\dfrac{\theta_1 - \gamma\theta_2}{(1-\gamma)(\theta_1 + \theta_2)} C\right) - (\theta_1 - \gamma\theta_2), & \text{if } \gamma\theta_2 \leq \theta_1 \leq \theta_2/\gamma; \\ \\ U_r(C) - (\theta_1 - \gamma\theta_2), & \text{if } \theta_1 > \theta_2/\gamma. \end{cases} \tag{5.44}$$

From the preceding expression, we see that at a Nash equilibrium we would never have $\theta_1 > \theta_2/\gamma$, since in this case infinitesimally reducing θ_1 is a profitable deviation for user 1. Thus if $\boldsymbol{\theta}$ is a Nash equilibrium, then $\gamma\theta_1 \leq \theta_2$. Symmetrically, by reasoning for user 2, we see that $\gamma\theta_2 \leq \theta_1$.

Thus, in searching for Nash equilibria, we restrict our search to strategy vectors

(θ_1, θ_2) such that $\gamma\theta_1 \le \theta_2$ and $\gamma\theta_2 \le \theta_1$. This is actually the key insight of the proof, because under these restrictions the payoffs of each player become concave. This follows from (5.44): note that when $\gamma\theta_2 \le \theta_1 \le \theta_2/\gamma$, the payoff to player 1 is concave in θ_1.

Following this observation, the remainder of the proof closely follows the proof of Theorem 2.2, so we only sketch the argument. As in that proof, it is straightforward to show that at a Nash equilibrium, we must have $\theta_1 > 0$ and $\theta_2 > 0$. Furthermore, by differentiating (5.44) when $\gamma\theta_2 \le \theta_1 \le \theta_2/\gamma$, we find that if a pair (θ_1, θ_2) is a Nash equilibrium then the following conditions must hold for user 1, where $\mu = (1-\gamma)(\theta_1 + \theta_2)/C > 0$, and $d_1 = (\theta_1 - \gamma\theta_2)/\mu$:

$$\gamma U_1'(C) \ge \mu, \quad \text{if } d_1 = C; \tag{5.45}$$

$$U_1'(d_1)\left(1 - \frac{(1-\gamma)d_1}{C}\right) = \mu, \quad \text{if } 0 < d_1 < C; \tag{5.46}$$

$$U_1'(0) \le \mu, \quad \text{if } d_1 = 0. \tag{5.47}$$

The first condition corresponds to the case where $\theta_1 = \theta_2/\gamma$; the second condition corresponds to the case where $\gamma\theta_2 < \theta_1 < \theta_2/\gamma$; and the third condition corresponds to the case where $\theta_1 = \gamma\theta_2$. Conditions symmetric to (5.45)-(5.47) hold for user 2, with $d_2 = (\theta_2 - \gamma\theta_1)/\mu$. Conversely, suppose we are given (d_1, d_2) and $\mu > 0$ such that $d_1 + d_2 = C$, the conditions (5.45)-(5.47) hold for user 1, and their symmetric analogues hold for player 2. Define (θ_1, θ_2) according to:

$$\theta_1 = \frac{\mu d_1 + \gamma\mu d_2}{1-\gamma^2}; \quad \theta_2 = \frac{\mu d_2 + \gamma\mu d_1}{1-\gamma^2}.$$

Then it follows that $(1-\gamma)(\theta_1+\theta_2)/C = \mu$, $d_1 = (\theta_1 - \gamma\theta_2)/\mu$, and $d_2 = (\theta_2 - \gamma\theta_1)/\mu$. Furthermore, we have $\gamma\theta_1 \le \theta_2$ and $\gamma\theta_2 \le \theta_1$; and since the payoff $Q_r(\theta_r; \boldsymbol{\theta}_{-r})$ is concave for $r = 1, 2$ in this case, the conditions (5.45)-(5.47) (for both users 1 and 2) are sufficient to guarantee that $\boldsymbol{\theta}$ is a Nash equilibrium.

By arguing as in the proof of Theorem 2.2, it is straightforward to show from the optimality conditions (5.45)-(5.47) that the allocation (d_1, d_2) at any Nash equilibrium is the unique optimal solution to the following problem:

$$\text{maximize} \quad \hat{U}_1(d_1) + \hat{U}_2(d_2) \tag{5.48}$$

$$\text{subject to} \quad d_1 + d_2 = C; \tag{5.49}$$

$$d_1, d_2 \ge 0, \tag{5.50}$$

where for $r = 1, 2$,

$$\hat{U}_r(d_r) = \left(1 - \frac{(1-\gamma)d_r}{C}\right) U_r(d_r) + \left(\frac{(1-\gamma)d_r}{C}\right)\left(\frac{1}{d_r}\int_0^{d_r} U_r(z)\, dz\right). \tag{5.51}$$

(Note that for $\gamma = 0$, this definition is the same as (2.19).) As $\gamma \to 1$, we have $\hat{U}_r \to U_r$ pointwise, and hence uniformly, over $[0, C]$. It then immediately follows that the optimal solution to (5.48)-(5.50) (i.e., the Nash equilibrium allocation) converges to the optimal solution to $\mathit{SYSTEM}(C, 2, \mathbf{U})$, which establishes the desired result. □

The key step in the previous proof involves showing that we can restrict attention to strategy vectors where $\gamma\theta_1 \leq \theta_2$ and $\gamma\theta_2 \leq \theta_1$, and in this case, the payoff to both players becomes concave. Unfortunately, this step is critically dependent on the assumption that we have only two players. Thus an interesting question concerns extending the preceding result to a general setting of R players.

■ 5.2 Multiple Producers, Inelastic Demand

In this section, we turn our attention to a game where demand is inelastic, rather than supply; and where multiple producers compete to meet demand. The model we consider is motivated by the setting of Section 4.1. Let D denote the inelastic demand; we assume the demand is infinitely divisible among N producers, where $N > 1$. As in the previous section, our goal is to characterize smooth market-clearing mechanisms which minimize efficiency loss. Note that by contrast to the previous section, however, we fix N *a priori*; thus the mechanisms we consider may depend on N. Our main motivation for this shift is the difference between the analyses of Section 2.2 and Section 4.2. In Section 2.2, we found the worst case efficiency loss occurred as the number of users R approached infinity; for this reason, in Section 5.1, we considered efficiency loss of smooth market-clearing mechanisms independent of the value of R. By contrast, in Section 4.2 we found that the efficiency loss approached zero as the number of firms approached infinity. Thus, to develop a meaningful characterization theorem in this section, and in particular to make a meaningful comparison to the result of Theorem 4.3, we fix the number of firms N in advance. (We will find, however, that the efficiency loss minimizing mechanism does not depend on N.)

We begin by describing the producers, or firms. Each producer has a cost function $C : \mathbb{R} \to \mathbb{R}^+$. We assume that C is continuous on $\mathbb{R}$, and strictly increasing and convex on $\mathbb{R}^+$, with $C(s) = 0$ for $s \leq 0$; note that these are the same conditions as Assumption

4.1. Let $\mathcal{C}$ denote the set of possible cost functions; i.e.:

$$\mathcal{C} = \{C : \mathbb{R} \to \mathbb{R}^+ \mid C \text{ is continuous on } \mathbb{R}, \text{ and strictly increasing and convex on } \mathbb{R}^+, \\ \text{and } C(s) = 0 \text{ for } s \leq 0\}.$$

Let $\mathbf{C} = (C_1, \ldots, C_N)$ denote the vector of cost functions, where C_n is the cost function of firm n. We call a vector $\mathbf{C} \in \mathcal{C}^N$ a *cost system*; our goal will be to design a resource allocation mechanisms which is efficient for all cost systems.

We assume that cost is measured in monetary units; thus, if firm n produces an amount s_n, but receives revenue w_n, the net payoff to firm n is given by:

$$w_n - C_n(s_n).$$

Given any cost system $\mathbf{C}$, our goal is to minimize aggregate production cost, as defined in the following problem:

SYSTEM$(D, N, \mathbf{C})$:

$$\text{minimize} \quad \sum_{n=1}^{N} C_n(s_n) \tag{5.52}$$

$$\text{subject to} \quad \sum_{n=1}^{N} s_n = D; \tag{5.53}$$

$$\mathbf{s} \geq 0. \tag{5.54}$$

We will say that s *solves SYSTEM*$(D, N, \mathbf{C})$ if s is an optimal solution to (5.52)-(5.54), given the cost system $\mathbf{C}$.

As in the previous section, we will consider smooth market-clearing mechanisms to determine the production allocation. We have the following definition.

Definition 5.4
Given $D > 0$ and $N > 1$, a smooth market-clearing mechanism for D and N *is a differentiable function $S : (0, \infty) \times [0, \infty) \to \mathbb{R}^+$ such that for all nonzero $\boldsymbol{\theta} \in (\mathbb{R}^+)^N$, there exists a unique solution $p > 0$ to the following equation:*

$$\sum_{n=1}^{N} S(p, \theta_n) = D.$$

We let $p_S(\boldsymbol{\theta})$ denote this solution.

As in Definition 5.1, note that while this definition implicitly restricts the strategy θ_n of each firm to $\mathbb{R}^+$, this fact is inessential; the subsequent analysis can be adapted to hold even if the strategy space of each firm is allowed to be $[c, \infty)$, where $c \in \mathbb{R}$. We also observe that the market-clearing price is undefined if $\boldsymbol{\theta} = \mathbf{0}$. For this reason, in defining the game associated with a given mechanism S, we will assume that the payoff to all firms is $-\infty$ if the composite strategy is $\mathbf{0}$, analogous to the definition made in Section 5.1. In fact, we note that at this level, Definition 5.4 is identical to Definition 5.1, except that the mechanism may depend on both D and N.

We also observe that as in the previous section, this definition of smooth market-clearing mechanism generalizes the supply function interpretation of the mechanism discussed in Section 4.1. We recall that in that development, each firm submits a supply function of the form $S(p, \theta) = D - \theta/p$, and the resource manager chooses a price $p_S(\boldsymbol{\theta})$ to ensure that $\sum_{n=1}^{N} S(p, \theta_n) = D$. Thus, for this mechanism, we have $p_S(\boldsymbol{\theta}) = \sum_{n=1}^{N} \theta_n/((N-1)D)$ if $\boldsymbol{\theta} \neq 0$. Another possible mechanism is given by $S(p, \theta) = D - \theta/\sqrt{p}$; it is straightforward to verify that $p_S(\boldsymbol{\theta}) = [\sum_{n=1}^{N} \theta_n/((N-1)D)]^2$ if $\boldsymbol{\theta} \neq 0$.

We will restrict attention to a particular class of smooth market-clearing mechanisms for C and N, which we define as follows.

Definition 5.5
Given $D > 0$ and $N > 1$, the class $\mathcal{S}(D, N)$ consists of all smooth market-clearing mechanism S for D and N such that:

1. *For all $C \in \mathcal{C}$, a firm's payoff is concave if the firm is price taking; that is, for all $p > 0$ the function:*
$$pS(p, \theta) - C(S(p, \theta))$$
is concave for $\theta \geq 0$.

2. *For all $C_n \in \mathcal{C}$, a firm's payoff is concave if the firm is price anticipating; that is, for all $\boldsymbol{\theta}_{-n} \in (\mathbb{R}^+)^{N-1}$, the function:*
$$p_S(\boldsymbol{\theta})S(p_S(\boldsymbol{\theta}), \theta_n) - C_n(S(p_S(\boldsymbol{\theta}), \theta_n))$$
is concave in $\theta_n > 0$ if $\boldsymbol{\theta}_{-n} = \mathbf{0}$, and concave in $\theta_n \geq 0$ if $\boldsymbol{\theta}_{-n} \neq \mathbf{0}$.

3. *The function S is uniformly less than or equal to D; i.e., for all $p > 0$ and $\theta \geq 0$, $S(p, \theta) \leq D$.*

Our justification for these conditions is similar to the discussion following Definition 5.2 in Section 5.1.2. The only modification is Condition 3, where we require the supply functions to be bounded below D; this is a natural assumption, since the demand is known in advance and we do not expect any producer to supply more than the demand.

In order to state the main result, we define *competitive equilibrium* and *Nash equilibrium*. Given a cost system $\mathbf{C}$ and a smooth market-clearing mechanism $S \in \mathcal{S}(D, N)$, we say that a nonzero vector $\boldsymbol{\theta} \in (\mathbb{R}^+)^N$ is a competitive equilibrium if, for $\mu = p_S(\boldsymbol{\theta})$, there holds for all n:

$$\theta_n \in \arg\max_{\overline{\theta}_n \geq 0} \left[\mu S(\mu, \overline{\theta}_n) - C_n(S(\mu, \overline{\theta}_n))\right]. \tag{5.55}$$

Similarly, we say that a nonzero vector $\boldsymbol{\theta} \in (\mathbb{R}^+)^N$ is a Nash equilibrium if there holds for all n:

$$\theta_n \in \arg\max_{\overline{\theta}_n \geq 0} P_n(\overline{\theta}_n; \boldsymbol{\theta}_{-n}), \tag{5.56}$$

where

$$P_n(\theta_n; \boldsymbol{\theta}_{-n}) = \begin{cases} p_S(\boldsymbol{\theta}) S(p_S(\boldsymbol{\theta}), \theta_n) - C_n(S(p_S(\boldsymbol{\theta}), \theta_n)), & \text{if } \boldsymbol{\theta} \neq 0; \\ -\infty, & \text{if } \boldsymbol{\theta} = 0. \end{cases} \tag{5.57}$$

Notice that the payoff is $-\infty$ if the composite strategy vector is $\boldsymbol{\theta} = \mathbf{0}$, since in this case no market-clearing price exists.

Our interest is in the worst-case ratio of aggregate cost at any Nash equilibrium to the optimal value of *SYSTEM*$(D, N, \mathbf{C})$. Formally, for $S \in \mathcal{S}(D, N)$ we define a constant $\rho(D, N, S)$ as follows:

$$\rho(D, N, S) = \sup \left\{ \frac{\sum_{n=1}^N C_n(S(p_S(\boldsymbol{\theta}), \theta_n))}{\sum_{n=1}^N C_n(s_n)} \,\middle|\, \mathbf{C} \in \mathcal{C}^N, \mathbf{s} \text{ solves } \mathit{SYSTEM}(D, N, \mathbf{C}) \text{ and } \boldsymbol{\theta} \text{ is a Nash equilibrium} \right\}$$

In this case we define ρ as a function of the demand level D, the number of firms N, and the mechanism S. As previewed in the discussion above, this allows us to compare different market mechanisms for fixed values of N, and hence explore the relationship to the results of Chapter 4. Note that since all $C \in \mathcal{C}$ are strictly increasing and nonnegative, and $D > 0$, $\sum_{n=1}^N C_n(s_n)$ is strictly positive for any cost system $\mathbf{C}$ and any optimal solution $\mathbf{s}$ to *SYSTEM*$(D, N, \mathbf{C})$. However, Nash equilibria may not exist for some cost systems $\mathbf{C}$; in this case we set $\rho(D, N, S) = \infty$.

The following theorem shows that among smooth market-clearing mechanisms for which there always exists a fully efficient competitive equilibrium, the mechanism proposed in Section 4.1 minimizes efficiency loss when users are price anticipating.

Theorem 5.9

Assume $D > 0$ and $N > 1$. Let $S \in \mathcal{S}(D, N)$ be a smooth market-clearing mechanism for D and N such that for all cost systems $\mathbf{C} \in \mathcal{C}^N$, there exists a competitive equilibrium $\boldsymbol{\theta}$ such that $(S(p_S(\boldsymbol{\theta}), \theta_n), n = 1, \ldots, N)$ solves SYSTEM$(D, N, \mathbf{C})$. Then:

1. *$\rho(D, 2, S) = \infty$.*

2. *If $N > 2$, then there exists a concave, strictly increasing, differentiable, and invertible function $B : (0, \infty) \to (0, \infty)$ such that for all $p > 0$ and $\theta \geq 0$:*

$$S(p, \theta) = D - \frac{\theta}{B(p)}.$$

3. *For $N > 2$, $\rho(D, N, S) \geq 1 + 1/(N - 2)$, and this bound is met with equality if and only if $S(p, \theta) = D - \Delta\theta/p$ for some $\Delta > 0$.*

Proof. The proof proceeds as follows. We first use Condition 1 in Definition 5.5 to show that any mechanism $S \in \mathcal{S}(D, N)$ must be of the form $S(p, \theta) = a(p) - b(p)\theta$; this result is analogous to the start of the proof of Theorem 5.1. We then show that $a(p) = D$, and $b(p) > 0$; thus $S(p, \theta) = D - b(p)\theta$. Finally, we explicitly determine conditions that must be satisfied by $B(p) = 1/b(p)$, and compute the worst case efficiency loss for any mechanism satisfying these conditions.

We begin with the following lemma, which is an analogue of Lemma 5.2.

Lemma 5.10 *Let S be a smooth market-clearing mechanism for D and N. Then $S \in \mathcal{S}(D, N)$ if and only if the following three properties hold:*

1. *There exist functions $a, b : (0, \infty) \to \mathbb{R}$ such that for all $p > 0$ and $\theta \geq 0$, $S(p, \theta) = a(p) - b(p)\theta$. Furthermore, $a(p) \leq D$, and $b(p) \geq 0$.*

2. *For all $\boldsymbol{\theta}_{-n} \in (\mathbb{R}^+)^{N-1}$, the functions $-S(p_S(\boldsymbol{\theta}), \theta_n)$ and $p_S(\boldsymbol{\theta})S(p_S(\boldsymbol{\theta}), \theta_n)$ are concave in $\theta_n > 0$ if $\boldsymbol{\theta}_{-n} = \mathbf{0}$, and concave in $\theta_n \geq 0$ if $\boldsymbol{\theta}_{-n} \neq \mathbf{0}$.*

Proof. The proof is identical to the proof of Lemma 5.2; i.e., we consider a linear cost function $C(s) = \alpha s$, and then consider limits as $\alpha \to 0$ and $\alpha \to \infty$. The only additional subtlety involves showing that $a(p) \leq D$ and $b(p) \geq 0$. If $a(p) > D$, then by choosing $\theta = 0$ we have $S(p, \theta) > D$, violating Condition 3 in Definition 5.5; thus $a(p) \leq D$. On the other hand, suppose that $b(p) < 0$ for some p. Then by choosing θ large enough, we would have $S(p, \theta) > D$, violating Condition 3 in Definition 5.5. Thus $b(p) \geq 0$. □

We now show that if $S \in \mathcal{S}(D, N)$ satisfies the conditions of the theorem, then $a(p) = D$. Suppose not; then $a(p) < D$. Choose a vector $\mathbf{s}$ such that $\sum_{n=1}^{N} s_n = D$; $a(p) < s_1 < D$; $s_n > 0$ for all n; and $s_n \neq s_m$ for $m \neq n$. Next, choose a collection of strictly convex, differentiable cost functions $(C_1, \ldots, C_N)$ such that $C_n'(s_n) = p$. Then it is straightforward to establish that the unique optimal solution to *SYSTEM*$(D, N, \mathbf{C})$ is the vector $\mathbf{s}$.

Now let $\boldsymbol{\theta}$ be a competitive equilibrium such that the resulting allocation solves *SYSTEM*$(D, N, \mathbf{C})$; then we must have $S(p_S(\boldsymbol{\theta}), \theta_n) = s_n$ for all n. Since $s_n \neq s_m$

for $n \neq m$ and $N > 1$, it follows that $b(p_S(\boldsymbol{\theta})) > 0$, and $\theta_n > 0$ for at least one n. Differentiating (5.55) with $\mu = p_S(\boldsymbol{\theta})$, we find that:

$$C_n'(S(\mu, \theta_n)) \cdot \frac{\partial S(\mu, \theta_n)}{\partial \theta_n} = \mu \cdot \frac{\partial S(\mu, \theta_n)}{\partial \theta_n}.$$

But note that $\partial S(\mu, \theta_n)/\partial \theta_n = b(\mu) > 0$, so that the preceding relation reduces to:

$$C_n'(s_n) = C_n'(S(\mu, \theta_n)) = \mu = p_S(\boldsymbol{\theta}),$$

from which we conclude that $p_S(\boldsymbol{\theta}) = p$ (since C_n is strictly convex). But now consider firm 1: we have $a(p) - b(p)\theta_1 = s_1 \in (a(p), D)$. Since $b(p) > 0$, this is only possible if $\theta_1 < 0$. But we have assumed the strategy space of each firm is restricted to $[0, \infty)$, so this is a contradiction; thus we must have $a(p) = D$.

Thus any mechanism $S \in \mathcal{S}(D, N)$ satisfying the conditions of the theorem must be of the form:

$$S(p, \theta) = D - b(p)\theta, \quad p > 0,\ \theta \geq 0.$$

Observe that this immediately implies $b(p) > 0$ for all $p > 0$; otherwise, if $b(p) = 0$, we can argue as in the previous paragraph and construct a competitive equilibrium for which $p_S(\boldsymbol{\theta}) = p$, which would imply $S(p_S(\boldsymbol{\theta}), \theta_n) = D$ for all n. But then $\sum_{n=1}^N S(p_S(\boldsymbol{\theta}), \theta_n) = ND > D$ (since $N > 1$), a contradiction to the assumption that $p_S(\boldsymbol{\theta})$ is a market-clearing price.

Since $b(p) > 0$, we let $B(p) = 1/b(p)$; thus $S(p, \theta) = D - \theta/B(p)$. For fixed $N > 1$ and nonzero $\boldsymbol{\theta} \geq 0$, since $\sum_{n=1}^N S(p_S(\boldsymbol{\theta}), \theta_n) = D$, we have:

$$B(p_S(\boldsymbol{\theta})) = \frac{\sum_{n=1}^N \theta_n}{(N-1)D}. \tag{5.58}$$

Thus we find, for nonzero $\boldsymbol{\theta}$:

$$S(p_S(\boldsymbol{\theta}), \theta_n) = D - \left(\frac{\theta_n}{\sum_{m=1}^N \theta_m} \right) (N-1)D. \tag{5.59}$$

In particular, observe that the preceding expression is convex in θ_n, for nonzero $\boldsymbol{\theta}$.

We now claim that B is invertible on $(0, \infty)$. To see this, note from (5.58) that B is clearly onto, since the right hand side of (5.58) can take any value in $(0, \infty)$. Furthermore, if $B(p_1) = B(p_2) = \gamma$ for some prices $p_1, p_2 > 0$, then if we choose $\boldsymbol{\theta}$ such that $\sum_{n=1}^N \theta_n/((N-1)D) = \gamma$, we find that $p_S(\boldsymbol{\theta})$ is not uniquely defined. Thus B is one-to-one as well, and hence invertible. Finally, since S is differentiable, B must be differentiable as well.

We let Φ denote the differentiable inverse of B. Observe from Condition 2 in Lemma 5.10 that $w_n(\boldsymbol{\theta}) = p_S(\boldsymbol{\theta}) S(p_S(\boldsymbol{\theta}), \theta_n)$, the revenue to firm n, must be concave in

θ_n for nonzero $\boldsymbol{\theta}$. From (5.58) and (5.59), we have:

$$w_n(\boldsymbol{\theta}) = \Phi\left(\frac{\sum_{m=1}^N \theta_m}{(N-1)D}\right)\left(D - \frac{\theta_n}{\sum_{m=1}^N \theta_m}(N-1)D\right). \tag{5.60}$$

We will consider the cases $N = 2$ and $N > 2$ separately. Suppose first that $N = 2$. In this case we have:

$$w_1(\theta_1, \theta_2) = \Phi\left(\frac{\theta_1 + \theta_2}{D}\right)\left(\frac{\theta_2 D}{\theta_1 + \theta_2}\right).$$

(A symmetric expression holds for $w_2(\theta_1, \theta_2)$.) Now observe that the preceding quantity is nonnegative for every nonzero (θ_1, θ_2). Thus if $\theta_2 > 0$, then $w_1(\theta_1, \theta_2)$ is a concave, nonnegative function of $\theta_1 \geq 0$; this is only possible if $\partial w_1(\theta_1, \theta_2)/\partial \theta_1 \geq 0$ for all $\theta_1 \geq 0$.

Now suppose that $(C_1, C_2) \in \mathcal{C}^2$ is a cost system, and that $\boldsymbol{\theta}$ is a Nash equilibrium. By our definition of P_n in (5.57), at least one of θ_1 or θ_2 must be nonzero; assume without loss of generality that $\theta_2 > 0$. Now consider the payoff to firm 1. We have already shown that $w_1(\theta_1, \theta_2)$ is nondecreasing in θ_1. Furthermore, from (5.59), it is straightforward to show that $S(p_S(\boldsymbol{\theta}), \theta_1) = \theta_2 D/(\theta_1 + \theta_2)$. Thus, if $\theta_2 > 0$ then the cost $C_1(S(p_S(\boldsymbol{\theta}), \theta_1))$ is strictly decreasing as θ_1 increases (since C_1 is assumed to be strictly increasing). Thus the payoff to firm 1 is strictly improved by increasing θ_1, and $\boldsymbol{\theta}$ could not have been a Nash equilibrium. We conclude no Nash equilibrium exists, so that $\rho(D, 2, S) = \infty$. (Note that in fact, no Nash equilibrium exists for *any* cost system.) This establishes the first claim of the theorem.

For the remainder of the proof, then, we assume that $N > 2$. In this case, we will show that Φ is strictly increasing and convex. Since w_n must be concave in θ_n, it follows easily that Φ must be convex. If not, then by considering strategy vectors $\boldsymbol{\theta}$ where $\boldsymbol{\theta}_{-n} = 0$ and $\theta_n = (N-1)D\mu$, we can show that w_n is not concave in θ_n (since in this case we have $w_n(\boldsymbol{\theta}) = (2-N)D\Phi(\mu)$). Consequently, it only remains to be shown that Φ is strictly increasing. Since Φ is invertible, it must be strictly monotonic; and thus Φ is either strictly increasing or strictly decreasing. We differentiate (5.60) with respect to θ_n:

$$\frac{\partial w_n(\boldsymbol{\theta})}{\partial \theta_n} = \Phi'\left(\frac{\sum_{m=1}^N \theta_m}{(N-1)D}\right)\left(\frac{1}{N-1} - \frac{\theta_n}{\sum_{m=1}^N \theta_m}\right) - \Phi\left(\frac{\sum_{m=1}^N \theta_m}{(N-1)D}\right)\left(\frac{\sum_{m\neq n} \theta_m}{\left(\sum_{m=1}^N \theta_m\right)^2}\right)(N-1)D. \tag{5.61}$$

Suppose that Φ is strictly decreasing; then $\Phi' < 0$. Choose $\boldsymbol{\theta}$ such that $\theta_n/(\sum_{m=1}^N \theta_m) =$

$1/(N-1)$, and consider infinitesimally increasing θ_n. The first term in (5.61) becomes nonnegative (in particular, it does not decrease), while the second term strictly increases. Thus w_n could not be concave; we conclude that Φ must be strictly increasing.

We summarize these observations in the following lemma, which also establishes the second claim of the theorem.

Lemma 5.11 *Suppose $N > 2$. If a smooth market-clearing mechanism $S \in \mathcal{S}(D, N)$ satisfies the conditions of the theorem, then there exists a concave, strictly increasing, and differentiable function $B : (0, \infty) \to (0, \infty)$ such that $S(p, \theta) = D - \theta/B(p)$ for all $p > 0$ and $\theta \geq 0$. Furthermore, B is invertible, so that $B(p) \to 0$ as $p \to 0$ and $B(p) \to \infty$ as $p \to \infty$. Finally, for nonzero $\boldsymbol{\theta} \in (\mathbb{R}^+)^N$, there holds:*

$$p_S(\boldsymbol{\theta}) = \Phi\left(\frac{\sum_{n=1}^N \theta_n}{(N-1)D}\right),$$

where Φ is the inverse of B.

Proof of Lemma. Since B has already been shown to be invertible, and Φ has been shown to be strictly increasing and convex, it is clear that B is strictly increasing and concave, with $B(p) \to 0$ as $p \to 0$, and $B(p) \to \infty$ as $p \to \infty$. □

In contrast to the proof of Theorem 5.1 (and in particular Lemma 5.3), convexity of Φ is not a sufficient condition for S to be a member of $\mathcal{S}(D, N)$; and furthermore, Nash equilibria are not guaranteed to exist for all members of $\mathcal{S}(D, N)$ satisfying the conditions of the theorem. Nevertheless, the following lemma gives necessary and sufficient optimality conditions which characterize Nash equilibria.

Lemma 5.12 *Let $S \in \mathcal{S}(D, N)$ satisfy the conditions of the theorem, and let Φ be the inverse of B as given in Lemma 5.11. Let $\mathbf{C}$ be a cost system. A vector $\boldsymbol{\theta} \geq 0$ is a Nash equilibrium if and only if at least two components of $\boldsymbol{\theta}$ are nonzero, and there exists a nonzero vector $\mathbf{s} \geq 0$ and a scalar $\mu > 0$ such that $\theta_n = \mu(D - s_n)$ for all n, $\sum_{n=1}^N s_n = D$, and the following conditions hold:*

$$(\mu\Phi'(\mu) - \Phi(\mu))\left(\frac{s_n}{(N-2)D}\right) + \left(1 + \frac{s_n}{(N-2)D}\right)\frac{\partial^- C_n(s_n)}{\partial s_n} \leq \Phi(\mu), \quad \text{if } 0 < s_n \leq D; \tag{5.62}$$

$$(\mu\Phi'(\mu) - \Phi(\mu))\left(\frac{s_n}{(N-2)D}\right) + \left(1 + \frac{s_n}{(N-2)D}\right)\frac{\partial^+ C_n(s_n)}{\partial s_n} \geq \Phi(\mu), \quad \text{if } 0 \leq s_n < D. \tag{5.63}$$

In this case $s_n = S(p_S(\boldsymbol{\theta}), \theta_n)$, and $\Phi(\mu) = p_S(\boldsymbol{\theta})$.

Proof. First suppose that $\boldsymbol{\theta}$ is a Nash equilibrium. From (5.57), the payoff to all players is $-\infty$ if $\boldsymbol{\theta} = 0$, so at least one component of $\boldsymbol{\theta}$ is positive. Suppose then that $\boldsymbol{\theta}_{-n} = 0$ for some firm n, while $\theta_n > 0$. Then using (5.59) and (5.60), the payoff to this firm is:

$$P_n(\theta_n; \boldsymbol{\theta}_{-n}) = \Phi\left(\frac{\theta_n}{(N-1)D}\right)(2-N)D,$$

since $S(p_S(\boldsymbol{\theta}), \theta_n) = (2-N)D < 0$ implies $C_n(S(p_S(\boldsymbol{\theta}), \theta_n)) = 0$, by our assumptions on $C_n \in \mathcal{C}$. But now observe that firm n can improve its payoff by infinitesimally reducing θ_n; thus $\boldsymbol{\theta}$ cannot be a Nash equilibrium. We conclude that at a Nash equilibrium, at least two components of $\boldsymbol{\theta}$ must be positive.

Next, let $\mu = \sum_{m=1}^{N} \theta_m/((N-1)D) > 0$, and $s_n = S(p_S(\boldsymbol{\theta}), \theta_n) = D - \theta_n/\mu$ (where the last equality follows from (5.59)). Then observe that $\theta_n = \mu(D - s_n)$. We now show that s ≥ 0; suppose, to the contrary, that $s_n < 0$ for firm n. Then the payoff to firm n is $P_n(\theta_n; \boldsymbol{\theta}_{-n}) = \Phi(\mu)s_n < 0$, since $C_n(s_n) = 0$. On the other hand, by choosing $\overline{\theta}_n = (\sum_{m\neq n} \theta_m)/(N-2) > 0$, it follows from (5.59) that $S(p_S(\overline{\theta}_n, \boldsymbol{\theta}_{-n}), \overline{\theta}_n) = 0$, and thus $P_n(\overline{\theta}_n; \boldsymbol{\theta}_{-n}) = 0 > P_n(\theta_n; \boldsymbol{\theta}_{-n})$, so that $\boldsymbol{\theta}$ could not have been a Nash equilibrium. We conclude that at a Nash equilibrium, the production of every firm is nonnegative, i.e., s ≥ 0. Alternatively, this implies that $0 \leq \theta_n \leq (\sum_{m\neq n} \theta_m)/(N-2)$ for all n.

Finally, observe that when $\boldsymbol{\theta}_{-n}$ is nonzero, the payoff $P_n(\overline{\theta}_n; \boldsymbol{\theta}_{-n})$ to firm n is directionally differentiable for $\overline{\theta}_n \geq 0$; it is also concave, by Condition 2 in Definition 5.5. Thus necessary and sufficient conditions for (5.56) to hold are:

$$\frac{\partial^+ P_n(\theta_n; \boldsymbol{\theta}_{-n})}{\partial \theta_n} \leq 0, \quad \text{if } 0 \leq \theta_n < \frac{\sum_{m\neq n} \theta_m}{N-2}; \tag{5.64}$$

$$\frac{\partial^- P_n(\theta_n; \boldsymbol{\theta}_{-n})}{\partial \theta_n} \geq 0, \quad \text{if } 0 < \theta_n \leq \frac{\sum_{m\neq n} \theta_m}{N-2}. \tag{5.65}$$

Note that using (5.59) and (5.60), we have:

$$P_n(\theta_n; \boldsymbol{\theta}_{-n}) = \Phi\left(\frac{\sum_{m=1}^{N} \theta_m}{(N-1)D}\right)\left(D - \left(\frac{\theta_n}{\sum_{m=1}^{N} \theta_m}\right)(N-1)D\right) - C_n\left(D - \left(\frac{\theta_n}{\sum_{m=1}^{N} \theta_m}\right)(N-1)D\right).$$

Differentiating this expression and using the substitutions $\mu = \sum_{m=1}^{N} \theta_m/((N-1)D)$, $\theta_n/(\sum_{m=1}^{N} \theta_m) = (D - s_n)/((N-1)D)$, and $s_n = D - \theta_n(N-1)D/(\sum_{m=1}^{N} \theta_m)$, (5.64)-(5.65) become equivalent to (5.62)-(5.63).

On the other hand, suppose that we have found $\boldsymbol{\theta}$, s, and μ such that the conditions of the lemma are satisfied. In this case we simply reverse the argument above; since $P_n(\overline{\theta}_n; \boldsymbol{\theta}_{-n})$ is concave in $\overline{\theta}_n$ (Condition 2 in Definition 5.5), if at least two components of $\boldsymbol{\theta}$ are nonzero then the conditions (5.64)-(5.65) are necessary and sufficient for $\boldsymbol{\theta}$ to be a Nash equilibrium. Furthermore, if s ≥ 0, $\mu > 0$, $\theta_n = \mu s_n$, and $\sum_{n=1}^N s_n = D$, then it follows that $\mu = \sum_{n=1}^N \theta_n/((N-1)D)$, $\Phi(\mu) = p_S(\boldsymbol{\theta})$, and $s_n = S(p_S(\boldsymbol{\theta}), \theta_n)$. Under these identifications the conditions (5.64)-(5.65) become equivalent to (5.62)-(5.63), as required. □

We now turn our attention to computing $\rho(D, N, S)$ for $S \in \mathcal{S}(D, N)$. Note that, if there exist cost systems $\mathbf{C}$ for which no Nash equilibrium exists under S, then trivially $\rho(D, N, S) = \infty$. For the remainder of the proof, we assume without loss of generality, therefore, that Nash equilibria exist for any cost system $\mathbf{C}$.

For $\mu > 0$ we let $\Gamma(\mu) = \mu\Phi'(\mu) - \Phi(\mu)$. Since Φ is convex and strictly increasing with $\Phi(\mu) \to 0$ as $\mu \to 0$, we know $\Gamma(\mu) \geq 0$, and $\Gamma(\mu)$ is nondecreasing. Suppose that $\Gamma(\mu) = 0$ for all μ. In this case $\Phi(\mu) = \Delta\mu$ for some $\Delta > 0$, and thus $B(p) = p/\Delta$. It then follows from Theorem 4.3 in Chapter 4 that $\rho(D, N, S) = 1 + 1/(N-2)$. For the remainder of the proof, therefore, we assume that $\Gamma(\mu) > 0$ for some $\mu > 0$.

We will need the following lemma.

Lemma 5.13 *Fix $S \in \mathcal{S}(D, N)$, and let Φ be defined as in Lemma 5.11. Suppose there exists t such that $0 \leq t < D$ and for all $\mu > 0$,*

$$\left(\frac{t}{(N-2)D}\right)\Gamma(\mu) > \Phi(\mu). \tag{5.66}$$

Then $\rho(D, N, S) = \infty$.

Proof. Of course, (5.66) can only hold if $t > 0$, so we assume this for the remainder of the proof. We define a cost system $\mathbf{C}$ as follows. Choose δ such that $0 < \delta < 1$. Define $C_1(s_1)$ according to:

$$C_1(s_1) = \begin{cases} 0, & \text{if } s_1 \leq 0; \\ \delta s_1, & \text{if } 0 \leq s_1 \leq t; \\ s_1 - t + \delta t, & \text{if } s_1 \geq t. \end{cases}$$

Then note that $C_1 \in \mathcal{C}$. For $n = 2, \ldots, N$ and $s_n \geq 0$, let $C_n(s_n) = \alpha s_n$, where $\alpha > 1$; and let $C_n(s_n) = 0$ for $s_n \leq 0$. Thus $C_n \in \mathcal{C}$ as well. Furthermore, in this case the optimal value of *SYSTEM*$(D, N, \mathbf{C})$ is equal to $D - t + \delta t$ (the entire demand is allocated to firm 1).

Now suppose that $\boldsymbol{\theta}$ is a Nash equilibrium, and let s and μ be defined as in Lemma 5.12. Suppose that $s_1 \geq t$. Since $t > 0$, it follows from (5.62) that:

$$\left(\frac{t}{(N-2)D}\right)\Gamma(\mu) \leq \left(\frac{s_1}{(N-2)D}\right)\Gamma(\mu) + \left(1 + \frac{s_1}{(N-2)D}\right)\frac{\partial^- C_1(s_1)}{\partial s_1} \leq \Phi(\mu),$$

which contradicts (5.66). Thus $0 \le s_1 < t$. But in this case, notice the aggregate cost at the Nash equilibrium is given by $\delta s_1 + \alpha(D - s_1) \ge \alpha(D - t)$. This inequality holds independent of the value of $\alpha > 1$; thus, we can choose a sequence of cost systems $\mathbf{C}$ with $\alpha \to \infty$ such that the ratio of Nash equilibrium aggregate cost to the optimal value of $SYSTEM(D, N, \mathbf{C})$ is lower bounded by:

$$\frac{\alpha(D-t)}{D-t+\delta t}.$$

As $\alpha \to \infty$, this ratio approaches ∞, which establishes that $\rho(D, N, S) = \infty$. □

In view of the previous lemma, we can restrict attention to mechanisms $S \in \mathcal{S}(D, N)$ such that for all t with $0 \le t < D$, there exists a scalar $\mu_t > 0$ such that:

$$\left(\frac{t}{(N-2)D}\right)\Gamma(\mu_t) \le \Phi(\mu_t).$$

If necessary, by replacing μ_t with $\mu_{t+\varepsilon}$ for sufficiently small $\varepsilon > 0$, we can make the following stronger assumption without loss of generality:

$$\left(\frac{t}{(N-2)D}\right)\Gamma(\mu_t) < \Phi(\mu_t). \tag{5.67}$$

Let t_0 be the smallest value of $t < D$ such that $t \ge D/N$ and $(D-t)^2/t \le 1$. For the remainder of the proof, we restrict attention to $t \in (t_0, D)$. Observe that for any such t we will have $(D-t)^2/t < 1$.

We now define a sequence of cost systems $\mathbf{C}^t$ as follows. Given t such that $t_0 < t < D$, choose $A_t > 0$ such that:

$$\left(\frac{t}{(N-2)D}\right)\Gamma(\mu_t) + A_t\left(1 + \frac{t}{(N-2)D}\right) = \Phi(\mu_t). \tag{5.68}$$

Define $\delta_t = (D-t)^2 A_t/t < A_t$, and let $C_1^t(s_1)$ be defined by:

$$C_1^t(s_1) = \begin{cases} 0, & \text{if } s_1 \le 0; \\ \delta_t s_1, & \text{if } 0 \le s_1 \le t; \\ A_t(s_1 - t) + \delta_t t, & \text{if } s_1 \ge t. \end{cases} \tag{5.69}$$

Since $0 < \delta_t < A_t$, we have $C_1^t \in \mathcal{C}$. Next, define α_t as:

$$\alpha_t = \left(1 + \frac{D-t}{(N-1)(N-2)D}\right)^{-1}\left(\Phi(\mu_t) - \left(\frac{D-t}{(N-1)(N-2)D}\right)\Gamma(\mu_t)\right). \tag{5.70}$$

Note that since we have assumed $t > t_0 \ge D/N$, we have $(D-t)/((N-1)(N-2)D) <$

$t/((N-2)D)$; thus applying (5.67) and (5.68), it follows that $\alpha_t > A_t$. For $n = 2, \dots, N$, define $C_n^t(s_n)$ as:

$$C_n^t(s_n) = \begin{cases} 0, & \text{if } s_n \le 0; \\ \alpha_t s_n, & \text{if } s_n \ge 0. \end{cases} \tag{5.71}$$

Since $\alpha_t > A_t > 0$, we have $C_n^t \in \mathcal{C}$. We now observe that the optimal solution to *SYSTEM*$(D, N, \mathbf{C}^t)$ is to allocate the entire demand to firm 1, which yields optimal value $\delta_t t + A_t(D-t)$.

Let $s_1 = t$, $s_n = (D-t)/(N-1)$ for $n = 2, \dots, N$, and $\mu = \mu_t$. Then note that $\sum_{n=1}^N s_n = D$, and $0 < s_n < D$ for all n. In particular, if we define $\theta_n = \mu(D - s_n)$ for all n, then $\mathbf{s}$, $\boldsymbol{\theta}$, and μ will satisfy the conditions of Lemma 5.12 as long as (5.62)-(5.63) are satisfied, in which case $\boldsymbol{\theta}$ would be a Nash equilibrium.

We now proceed to check that (5.62)-(5.63) are satisfied. Note that $\partial^+ C_1^t(s_1)/\partial s_1 = A_t$, and $\partial^- C_1^t(s_1)/\partial s_1 = \delta_t$. Thus (5.63) is satisfied with equality for firm 1 (by (5.68)), while $\delta_t < A_t$ implies that (5.62) holds as well. For the remaining users, we have $\partial C_n^t(s_n)/\partial s_n = \alpha_t$, and thus the definition of α_t in (5.70) ensures that (5.62)-(5.63) hold with equality. We conclude that $\boldsymbol{\theta}$ is a Nash equilibrium, and $s_n = S(p_S(\boldsymbol{\theta}), \theta_n)$ is the resulting allocation to firm n.

After substitution of (5.68) into (5.70), the ratio of Nash equilibrium aggregate cost $\sum_{n=1}^N C_n^t(s_n)$ to the optimal value of *SYSTEM*$(D, N, \mathbf{C}^t)$ is given by:

$$F(t) = \left(\frac{1}{\delta_t t + A_t(D-t)}\right) \cdot \left(\delta_t t + \frac{A_t\left(1 + \dfrac{t}{(N-2)D}\right) + \Gamma(\mu_t)\left(\dfrac{t}{(N-2)D} - \dfrac{D-t}{(N-1)(N-2)D}\right)}{1 + \dfrac{D-t}{(N-1)(N-2)D}}(D-t)\right).$$

If we now substitute for δ_t (recalling $\delta_t = (D-t)^2 A_t/t$), and normalize by $A_t(D-t)$, we have:

$$F(t) = \left(\frac{1}{D-t+1}\right) \cdot \left(D - t + \frac{\left(1 + \dfrac{t}{(N-2)D}\right) + \left(\dfrac{\Gamma(\mu_t)}{A_t}\right)\left(\dfrac{t}{(N-2)D} - \dfrac{D-t}{(N-1)(N-2)D}\right)}{1 + \dfrac{D-t}{(N-1)(N-2)D}}\right). \tag{5.72}$$

By taking subsequences if necessary, we assume without loss of generality that we have a sequence $t_k \to D$ such that the nonnegative sequences $\Gamma(\mu_{t_k})$, A_{t_k}, $\Phi(\mu_{t_k})$, $\Gamma(\mu_{t_k})/A_{t_k}$, and $\Phi(\mu_{t_k})/A_{t_k}$ all converge (possibly to ∞).

We will distinguish two cases. First suppose there exists $\mu > 0$ such that $\Phi(\mu) >$

$\Gamma(\mu)$. Since $\Phi(\mu)$ is strictly increasing, $\Gamma(\mu)$ is nondecreasing, and $\Gamma(\mu)$ is not identically zero, it can be verified that we can assume $\Gamma(\mu) > 0$ without loss of generality. In this case it follows from (5.67) we may let $\mu_t = \mu$ for all $t \in (t_0, D)$. With this definition of μ_t we can rewrite (5.68) as:

$$\left(\frac{t}{(N-2)D}\right)\left(\frac{\Gamma(\mu)}{A_t}\right) + \left(1 + \frac{t}{(N-2)D}\right) = \frac{\Phi(\mu)}{A_t}.$$

Since $t_k \to D$ as $k \to \infty$, it follows from this expression that we must have $A = \lim_{k\to\infty} A_{t_k} < \infty$, which implies:

$$\lim_{k\to\infty} F(t_k) = 1 + \frac{1}{N-2} + \left(\frac{1}{N-2}\right)\left(\frac{\Gamma(\mu)}{A}\right) > 1 + \frac{1}{N-2}.$$

(Of course, if $A = 0$, then $\lim_{k\to\infty} F(t_k) = \infty$.) This establishes that $\rho(D, N, S) > 1 + 1/(N-2)$ in this case. Note this depends critically on the fact that Γ is not identically zero; if $\Gamma(\mu) = 0$ for all μ, then this bound holds with equality.

For the remainder of the proof, therefore, we assume that for all $\mu > 0$, we have $\Phi(\mu) \le \Gamma(\mu)$. Let $\zeta = \lim_{k\to\infty} \Gamma(\mu_{t_k})/A_{t_k}$. Rewrite (5.68) as:

$$\left(\frac{t}{(N-2)D}\right)\left(\frac{\Gamma(\mu_t)}{A_t}\right) + \left(1 + \frac{t}{(N-2)D}\right) = \frac{\Phi(\mu_t)}{A_t}.$$

From the preceding expression, it follows that if $\zeta = 0$, then $\Phi(\mu_{t_k})/A_{t_k} \to 1 + 1/(N-2) > 0$ as $k \to \infty$ (since $t_k \to D$). But then we have:

$$\lim_{k\to\infty} \frac{\Phi(\mu_{t_k}) - \Gamma(\mu_{t_k})}{A_{t_k}} = 1 + \frac{1}{N-2} > 0,$$

so that for sufficiently large k we have $\Gamma(\mu_{t_k}) < \Phi(\mu_{t_k})$—a contradiction. Thus we must have $\zeta > 0$, in which case we have:

$$\lim_{k\to\infty} F(t_k) = 1 + \frac{1}{N-2} + \frac{\zeta}{N-2} > 1 + \frac{1}{N-2}.$$

(Of course, if $\zeta = \infty$, then $\lim_{k\to\infty} F(t_k) = \infty$.) This bound establishes that $\rho(D, N, S) > 1 + 1/(N-2)$, as required.

To summarize: we have shown that if $N > 2$ and $S(p, \theta) = D - \Delta\theta/p$ for $p > 0$, $\theta \ge 0$, then $\rho(D, N, S) = 1 + 1/(N-2)$ for $N > 2$. If $N > 2$, but there exists a $t \in [0, D)$ such that (5.66) holds, then $\rho(D, N, S) = \infty$. Finally, for all other $S \in \mathcal{S}(D, N)$ satisfying the conditions of the theorem, there holds $\rho(D, N, S) > 1 + 1/(N-2)$. This completes the proof of the third claim of the theorem. □

The preceding proof leaves two important open questions. First, unlike Lemma 5.3, we do not explicitly characterize necessary and sufficient conditions for a mechanism $S \in \mathcal{S}(D, N)$ to satisfy the conditions of the theorem. Furthermore, we do not establish that Nash equilibria always exist for such mechanisms. In particular, in light of these facts, it may be the case that no mechanisms in $\mathcal{S}(D, N)$ other than those of the form $S(p, \theta) = D - \Delta\theta/p$ satisfy the conditions of Theorem 5.9. We now demonstrate this is not the case, via the following example.

Example 5.1
Fix $N \geq 2$, and $D > 0$. Let $S(p, \theta) = D - \theta/\sqrt{p}$, for $p > 0$ and $\theta \geq 0$. Since $S(p, \theta) \leq D$, Condition 3 in Definition 5.5 is trivially satisfied. Furthermore, Condition 1 in Definition 5.5 is trivially satisfied, since $S(p, \theta)$ is linear in $\theta \geq 0$ for fixed $p > 0$. Finally, given a cost system $\mathbf{C}$, let $\mathbf{s}$ be an optimal solution to $\mathit{SYSTEM}(D, N, \mathbf{C})$, and let $p > 0$ be a Lagrange multiplier for the constraint (5.53). (Such a multiplier exists because the Slater constraint qualification holds; see [13].) Define $\theta = \sqrt{p}(D - s_n)$; then $S(p, \theta) = s_n$, and it is straightforward to verify that $\boldsymbol{\theta}$ is a competitive equilibrium, with $p_S(\boldsymbol{\theta}) = p$. Thus to check that S satisfies the conditions of Theorem 5.9, we only need to check Condition 2 in Definition 5.5.

We use Lemma 5.10. It is easy to check that:

$$p_S(\boldsymbol{\theta}) = \left(\frac{\sum_{m=1}^N \theta_m}{(N-1)D} \right)^2 .$$

As in (5.59), this implies:

$$S(p_S(\boldsymbol{\theta}), \theta_n) = D - \left(\frac{\theta_n}{\sum_{m=1}^N \theta_m} \right) (N-1)D.$$

Thus S is convex in θ_n, for nonzero $\boldsymbol{\theta}$. It remains to be checked that the expression $w_n(\boldsymbol{\theta}) = p_S(\boldsymbol{\theta}) S(p_S(\boldsymbol{\theta}), \theta_n)$ is concave in θ_n for nonzero $\boldsymbol{\theta}$, where:

$$w_n(\boldsymbol{\theta}) = \left(\frac{\sum_{m=1}^N \theta_m}{(N-1)D} \right)^2 \left(D - \frac{\theta_n}{\sum_{m=1}^N \theta_m}(N-1)D \right) = \frac{\left(\sum_{m=1}^N \theta_m\right)^2}{(N-1)^2 D} - \frac{\theta_n \left(\sum_{m=1}^N \theta_m\right)}{(N-1)D}.$$

It suffices to compute the coefficient on θ_n^2 in the expansion of $w_n(\boldsymbol{\theta})$; we have:

$$w_n(\boldsymbol{\theta}) = \frac{2-N}{(N-1)^2 D}\theta_n^2 + f(\boldsymbol{\theta}_{-n})\theta_n + g(\boldsymbol{\theta}_{-n}).$$

Here f and g are functions which depend only on the strategies of players other than firm n. Since $N \geq 2$, we conclude that w_n is concave in θ_n, as required. Thus $S \in \mathcal{S}(D, N)$.

We now show that if $N > 2$, then for any cost system $\mathbf{C}$, there always exists a Nash equilibrium under S. Fix a cost system $\mathbf{C}$; we will search for $\boldsymbol{\theta}$, $\mathbf{s}$, and μ which satisfy the conditions of Lemma 5.12. In the notation of the proof of Theorem 5.9, we have $S(p, \theta) = D - \theta/B(p)$, where $B(p) = \sqrt{p}$; thus $\Phi(\mu) = \mu^2$, and $\mu\Phi'(\mu) - \Phi(\mu) = \Phi(\mu) = \mu^2$. This allows us to rewrite the conditions (5.62)-(5.63) as follows:

$$\left(1 - \frac{s_n}{(N-2)D}\right)^{-1}\left(1 + \frac{s_n}{(N-2)D}\right)\frac{\partial^- C_n(s_n)}{\partial s_n} \leq \mu^2, \quad \text{if } 0 < s_n \leq D; \tag{5.73}$$

$$\left(1 - \frac{s_n}{(N-2)D}\right)^{-1}\left(1 + \frac{s_n}{(N-2)D}\right)\frac{\partial^+ C_n(s_n)}{\partial s_n} \geq \mu^2, \quad \text{if } 0 \leq s_n < D. \tag{5.74}$$

If $\mu^2 \leq \partial^+ C_n(0)/\partial s_n$, then let $s_n(\mu) = 0$; and if $(1 - 1/(N-2))^{-1}(1 + 1/(N-2))\partial^- C_n(D)/\partial s_n \leq \mu^2$, then let $s_n(\mu) = D$. For all other values of μ, (5.73)-(5.74) have a unique solution $s_n(\mu) \in (0, D)$. Furthermore, by arguing as in the proof of Lemma 5.5, it is straightforward to check that $s_n(\mu)$ is continuous and nondecreasing in μ, with $s_n(\mu) \to D$ as $\mu \to \infty$, and $s_n(\mu) \to 0$ as $\mu \to 0$. (This step uses the fact that the left hand sides of (5.73)-(5.74) are strictly increasing in s_n.) Thus there must exist $\mu > 0$ such that $\sum_n s_n(\mu) = D$. In this case the pair $\mathbf{s} = \mathbf{s}(\mu) \geq 0$ and $\mu > 0$ satisfy (5.12)-(5.13); and since $N > 2$ and $\mathbf{s} \geq 0$, at least two components of $\mathbf{s}$ must be strictly lower than D. Thus if we define $\theta_n = \mu(D - s_n)$, then the triple $\boldsymbol{\theta}$, $\mathbf{s}$, and $\mu > 0$ satisfies the conditions of Lemma 5.4, and hence $\boldsymbol{\theta}$ is a Nash equilibrium, as required. □

As in Section 5.1, a potentially undesirable feature of the mechanisms considered is that the payoff to firm n is defined as $-\infty$ when the composite strategy vector is $\boldsymbol{\theta} = 0$ (cf. (5.6)). We can restrict attention instead to mechanisms where $S(p, \theta) = D$ if $\theta = 0$, for all $p \geq 0$; in this case we can *define* $p_S(\boldsymbol{\theta}) = 0$ if $\boldsymbol{\theta} = 0$. From Lemma 5.11, if $p > 0$ and $\theta_n = 0$, then the payoff to firm n is $D - C_n(D)$; for the present discussion, then, we *redefine* the payoff to firm n to be $D - C_n(D)$ if $\theta_n = 0$, regardless of the strategy $\boldsymbol{\theta}_{-n}$ of the other firms. This condition amounts to a "normalization" on the market-clearing mechanism. Furthermore, this modification now captures the mechanism of Chapter 4, where $Q_r(0; \mathbf{w}_{-n}) = D - C_n(D)$ for all $\mathbf{w}_{-n} \geq 0$ (see (4.5), (4.6), and (4.16)). It is straightforward to check that this modification does not alter the conclusion of Theorem 5.9, since the class of mechanisms in $\mathcal{D}(C)$ satisfying the conditions of Theorem 5.9 can be extended to ensure $S(p, \theta) = D$ if $\theta = 0$.

For the mechanism discussed in Section 4.1, given by $S(p, \theta) = D - \theta/p$, another undesirable property is that $S(p, \theta)$ may be negative for some values of p and θ. This raises the following question: does there exist any mechanism $S \in \mathcal{S}(D, N)$ satisfying the conditions of Theorem 5.9, such that $S(p, \theta) \geq 0$ for all $p > 0$ and $\theta \geq 0$? The answer is no, from Lemma 5.11: we know that any such mechanism must have the form $S(p, \theta) = D - \theta/B(p)$, where $B(p) > 0$. Thus if we wish to design mechanisms where $S(p, \theta)$ is guaranteed to be nonnegative, we must relax either the definition of

$\mathcal{S}(D, N)$, or the assumptions in Theorem 5.1.

We conclude with a brief discussion of Condition 1 in Definition 5.5, the requirement that firms' payoffs be concave when they are price taking. The same objections to this assumption arise as in our discussion following Definition 5.5. However, by contrast to our analysis in Section 5.1.2, and in particular our definition of $\hat{\mathcal{D}}$ in that section, we cannot remove Condition 1 in Definition 5.5 by focusing instead on mechanisms which are well defined for all values of the inelastic demand D. The formal reason for this is simple: all the mechanisms in $\mathcal{S}(D, N)$ which satisfy the conditions of Theorem 5.9 explicitly depend on D, as shown in Lemma 5.11. Thus there remains an interesting open question: does the mechanism studied in Section 4.1 minimize efficiency loss among all smooth market-clearing mechanisms which satisfy only Conditions 2 and 3 of Definition 5.5 and the assumptions of Theorem 5.1? Part of the challenge in this problem lies in the fact that Lemma 5.10 no longer holds, and thus we do not have a simple characterization available of all such smooth market-clearing mechanisms.

■ 5.3 Chapter Summary

The results of this chapter characterize the mechanisms studied in Chapters 2 and 4 as those which uniquely satisfy certain desirable properties. These results are closely related in spirit to the classical literature on *mechanism design* in economics; we have framed and solved an axiomatic problem of developing efficient mechanisms given certain constraints (the most important being that the strategy spaces of the market participants should be one-dimensional). This relationship warrants further inspection, and in the following chapter we will situate our work in a broader mechanism design setting, and discuss related open questions.

CHAPTER 6

Conclusion

The central motif of this thesis is the investigation of efficiency loss in simple, market-clearing mechanisms for resource allocation, particularly for application in large scale systems. We concentrate on simple strategy spaces and single price mechanisms, in the belief that this simplicity yields scalability and robustness for distributed systems; on the other hand, we hope to ensure that even if market participants try to manipulate the market in their own self interest, efficiency losses will be bounded. This thesis represents one corner of a problem which requires a multifaceted approach; and in this chapter, we briefly discuss some key open issues on which further study is necessary: the modeling of two-sided markets (Section 6.1); dynamics (Section 6.2); and the relationship between classical mechanism design and modern distributed systems (Section 6.3). We conclude with final thoughts in Section 6.4.

■ 6.1 Two-Sided Markets

Throughout the thesis, we have concentrated only on competition occurring on one side of a market—either between consumers, or between producers. Of course, in practical large scale systems, competition typically occurs on *both* sides of the market simultaneously. Thus, for example, consumers may be competing for network resources, but in addition Internet service providers compete with each other to capture market share; and while generators compete to supply electricity, large buyers also compete to acquire electricity at the lowest price. If we view the architecture and infrastructure of large scale systems as guiding the interaction between market participants, then from an engineering perspective we must be sensitive to this competition between both buyers and sellers.

As a starting point, in our results we have ascribed higher precedence to one side of the market. In considering communication networks, the diversity and dispersion of network users suggests the approach of treating the problem of network resource allocation among users while taking the service provider's pricing strategy as fixed. Similarly, in considering electricity networks, the short run price inelasticity of demand suggests the approach of keeping demand fixed, while considering the competition among suppliers. An interesting open direction then concerns introducing into these models competition between both sides of the market, and investigating the resulting efficiency properties of the system as a whole.

6.2 Dynamics

This entire thesis considers only a *static* theory of equilibrium: either competitive equilibrium (when market participants are price taking), or Nash equilibrium (when market participants are price anticipating). These static equilibrium concepts only guarantee that if all market participants behave as prescribed by the equilibrium, then no player will have an incentive to deviate. This raises an obvious and critical question: how do market participants reach an equilibrium in the first place?

An answer to the previous question must, by definition, involve a *dynamic* model of market behavior. We have deliberately chosen the simpler route of considering static models in this thesis; but again, this only provides a starting point for a more substantive investigation of dynamic issues in network resource allocation problems. In this section we outline briefly some possible means for investigation of dynamics of the market mechanisms in this thesis.

Perhaps the simplest means of approaching dynamics is through the natural model of a *price-adjustment process* which matches supply and demand. We consider this approach here in the context of a single resource in inelastic supply, shared by multiple users (see Section 2.1). As shown in Theorem 2.2, at a Nash equilibrium it is *as if* market participants have achieved a competitive equilibrium, but with respect to "modified" utility functions. Once we make the connection between the Nash equilibria of a game where players are price anticipating and competitive equilibria of a related economic system, we can apply the well known dynamics of the *Walrasian tâtonnement process* [82, 137] to study the convergence of prices to a point where aggregate supply matches aggregate demand. (Indeed, these dynamics serve as inspiration for the dynamic models of Kelly et al. [65]; see also [127], as well as the introduction to Chapter 3, for further discussion and references.) Such dynamics immediately yield convergence to the Nash equilibrium for the simple setting of Section 2.1; and a similar investigation can be carried out for the more sophisticated models of this thesis.

However, the basic dynamics of the price-adjustment process are problematic, because they posit a very specific response on the part of market participants: prices move up if aggregate demand exceeds aggregate supply, and down otherwise; and the response of market participants is captured only through an immediate change in aggregate demand or supply in response to changes in price. However, in realistic scenarios it is unclear whether all players would react in the exact manner presupposed by the tâtonnement dynamics; for example, some players may react on faster timescales than others. Furthermore, players may not even be myopic in their response to the network state, and may use information from past interactions with the network to guide future behavior.

All these observations suggest that an applicable dynamic theory must allow for some degree of sophistication in the response of the players. Some modeling techniques include repeated games [43, 85], dynamic games [43], and learning in games [40, 42]. Incorporating elements of these approaches into the models considered here is appealing; however, dynamics in games are generally quite difficult to analyze, and for this reason an understanding of the dynamics of the market mechanisms presented here may be best achieved through experimental distributed game environments [67]. Indeed, one attraction of such experiments is that they investigate both the dynamics of users' behavior, as well as their interaction with network *software*, or protocols. The protocol stack controls the feedback to the users, as well as the strategies available to them; and thus limiting the users' actions through the protocol stack offers one possibility for mitigating the sophistication of users' responses.

The preceding discussion highlights the importance of *information* in models for games in large scale distributed systems. In order to adequately address the dynamic interaction of market participants, we must first understand the information available to them in making decisions. This information then guides both the responses of market participants, as well as the notion of equilibrium we choose. For this reason, we turn next to a discussion of the relationship between distributed systems, the models of this thesis, and the classical theory of *mechanism design*.

■ 6.3 Mechanism Design and Distributed Systems

In this section we will briefly discuss the connection of our work to the theory of *mechanism design*, and also frame some open questions regarding mechanism design for distributed systems. Mechanism design is a subfield of economics that seeks the *design* of games to achieve a prespecified outcome; typically, the theory is applied to achieve an efficient outcome in environments with multiple competing players. For an introductory survey of mechanism design, see Chapter 23 of [82].

In a resource allocation setting, mechanism design is concerned with the same basic problem we have treated in this thesis: how can resources be allocated efficiently among competing interests, even if players act in their own self interest? For simplicity, we consider here the basic framework of Section 2.1: multiple users compete for a scarce resource in inelastic supply with capacity C. Each user has a utility function describing his monetary value for the allocation that user receives; the goal is to allocate the scarce resource efficiently, i.e., to maximize aggregate utility.

We have considered the *Nash equilibrium* solution concept, which in full generality implies that users have complete information about their environment, including the utility functions of all other users. This is contrasted with implementation in *Bayesian* *equilibrium*, where users have some knowledge about other users, only through a probability distribution over the possible utility functions other users may have; a Bayesian equilibrium (or Bayes-Nash equilibrium) is then one where each user has chosen a strategy so as to maximize his expected payoff. Finally, in a *dominant strategy equilibrium*, users have no information about the utility functions of other users, and must choose strategies which maximize their payoffs independent of the choices of other players. Of course the dominant strategy implementation concept is the strongest concept, and dominant strategy equilibria are the least likely to exist.

These solution concepts have been extensively explored in the literature on mechanism design; we briefly survey some of the key contributions here. The seminal result of the field is the celebrated Vickrey-Clarke-Groves (VCG) mechanism [20, 50, 139], shown in quasilinear environments to be essentially the only mechanism which maximizes aggregate utility under the dominant strategy solution concept [46]. By contrast, there exist a plethora of such mechanisms when we expand to Bayesian implementation [27, 28, 4], and an interesting question concerns design of such mechanisms which maximize revenue to the resource seller [88]. However, implementation in Bayesian equilibrium (as well as Nash equilibrium) suffers from an unfortunate side effect: typically, there exist multiple equilibria [98], and only some of these may have desirable efficiency properties. (By contrast, under simple conditions the efficient outcome is the only dominant strategy equilibrium of the VCG mechanism [26, 75].) Finally, Moore has written a survey of implementation in Nash equilibrium [87]; the central result in this area is in the paper of Maskin [83], where conditions are given under which mechanisms can be designed that yield Pareto efficient Nash equilibria.

This literature almost exclusively focuses on the problem of finding *fully efficient* equilibria. The tradeoff for achieving full efficiency is typically a great increase in mechanism complexity. As an example, we outline the operation of the VCG mechanism. In such a mechanism, the strategy of a user is a “declared” utility function, which may or may not be the true utility function of that user. The VCG mechanism chooses an allocation which maximizes the aggregate declared utility of the users. The payment made by each user r is then the difference between the maximum possible aggregate utility of all other users if user r were not present, and the aggregate utility of all other users given that user r is present. (This payment rule is the “pivotal” rule [20].) Under these conditions, users have the incentive to declare their own utility function truthfully, independent of the strategy of other players [46].

The VCG mechanism is attractive because the notion of dominant strategy implementation implies users have the incentive to truthfully reveal preferences, without any knowledge of other users; this is a desirable feature in large networks, where users gain limited information about other users of the network. Notice, however, that the VCG mechanism has two undesirable features from the perspective of a large scale distributed system. First, the mechanism requires users to submit their entire utility function, a potentially high dimensional object. Second, the mechanism requires $R+1$ *centralized* computations of maximal aggregate utility (where R is the number of users). In general, it is difficult to design a computationally decentralized scheme, with low communications overhead, which can implement the VCG mechanism. Lazar and Semret [76, 116] have proposed a reduced dimension equivalent of the VCG mechanism for network resource allocation, and show that there exists a nearly fully efficient Nash equilibrium. However, their analysis does not rule out the possibility of other, potentially inefficient equilibria.

For this reason, we have chosen instead to focus instead on simple market-clearing mechanisms, and the design of such mechanisms to minimize efficiency loss (rather than guarantee full efficiency). The mechanisms we study have one-dimensional strategy spaces, and require only price feedback from the market. In particular, although we have considered Nash equilibria, in our model users need to know only the capacity C, their own strategy, and the current price to determine if they have made a payoff maximizing choice. Thus the mechanism of Section 2.1 appears to be better suited to a large scale architecture than the VCG mechanism.

We note, however, one caveat to the conclusion that our mechanism is "simple." The main objection is our use of Nash equilibrium as the solution concept, which requires "complete knowledge" of the utility functions. To help illustrate this point, consider the possibility of applying the Bayesian equilibrium concept instead. In this case the strategy of a player is a mapping from his realized utility function to a choice of bid. Thus, for a player to maximize expected payoff, he must choose a bid which is optimal (in expectation) relative to any possible bids that will be made by other players, given their choice of strategy. In particular, this means that at a Bayesian equilibrium, a player must choose a *single* bid which is an optimal response in an expectation taken over *multiple* possible prices. By contrast, because we adopt the Nash equilibrium solution concept in Section 2.1, a user's decision can be reduced to requiring knowledge only of the bids of other players, as well as the capacity. To summarize, the model of Section 2.1 does indeed yield the result that users only need knowledge of the capacity, the price, and their own strategy to determine whether they have made a payoff maximizing decision. However, this reduction in necessary information may also be interpreted as a *consequence* of the choice of the Nash equilibrium solution concept. (Indeed, a more complete resolution of this information issue requires a detailed understanding of a dynamic model for information acquisition and action by market participants.)

The field of mechanism design in distributed environments is only in its infancy, and the results of this thesis suggest the possibility of traditional market-clearing mechanisms as a solution. An alternate possibility is the adaptation of VCG and related mechanisms for use in distributed environments. This class of problems, broadly known as *distributed mechanism design*, has attracted attention in recent years, particularly in the computer science community; see, e.g., [39] for an overview.

We conclude our discussion of mechanism design with a note regarding power systems. While power systems are also large scale, distributed systems, they differ significantly from communication networks in that markets for electricity are run centrally at nodes throughout the power grid. As discussed in Chapter 4, these markets use *supply function bidding*, i.e., producers submit supply functions describing their willingness to produce electricity as a function of price. A market-clearing price is then chosen so supply equals demand.

Because these markets are run at central clearinghouses, a natural solution would be to implement exactly the VCG mechanism; indeed, one appealing feature of the VCG mechanism in this respect is that it is not affected by nonconvexity in the cost structure of producing firms. Hobbs et al. have presented an argument against the use of the VCG mechanism in electricity markets [56]; however, we believe that deeper investigation of the structure of the electricity market itself is necessary, to determine whether the goal of efficient allocation can be achieved using existing tools of mechanism design. By contrast, our approach in Chapter 4 is that the supply function bidding structure is a consequence of social and political realities which make single price, market-clearing mechanisms attractive, justifying our investigation of efficient market-clearing mechanisms (see also [147]).

■ 6.4 Future Directions

Any attempt to design efficient resource allocation mechanisms for large scale systems must answer at least the following questions:

1. What constitutes an efficient allocation?

2. What information is available to the market participants?

3. What constraints on mechanism complexity (regarding both computation and communication) are imposed by the system?

We have defined efficiency in terms of Pareto efficiency; we have assumed that market participants receive information from the mechanism in the form of a single market-clearing price, as well as an allocation of resources; and finally, we meet the requirements of simplicity of communication and computation through one-dimensional strategy spaces and a market-clearing process. However, as the discussion of this chapter has highlighted, these are all choices open to debate. Indeed, it is our hope that the results of this thesis serve as a springboard to further answers to the three questions posed above. In particular, we view the long term goal of such an agenda to be a more comprehensive view of the interaction between information, computation, and communication in large scale systems, and the impact of these components on mechanism design.

References

[1] E. J. Anderson and A. B. Philpott. Using supply functions for offering generation into an electricity market. *Operations Research*, 50(3):477–499, 2002. *(Cited on page 166.)*

[2] E. Anshelevich, A. Dasgupta, E. Tardos, and T. Wexler. Near-optimal network design with selfish agents. In *Proceedings of the 35th Annual ACM Symposium on the Theory of Computing*, pages 511–520, 2003. *(Cited on page 25.)*

[3] K. J. Arrow. *Social Choice and Individual Values*. John Wiley and Sons, Inc., New York, New York, 1951. *(Cited on page 19.)*

[4] K. J. Arrow. The property rights doctrine and demand revelation under incomplete information. In M. J. Boskin, editor, *Economics and Human Welfare: Essays in Honor of Tibor Scitovsky*, pages 23–39. Academic Press, New York, New York, 1979. *(Cited on page 234.)*

[5] K. J. Arrow and F. Hahn. *General Competitive Analysis*. Holden-Day, Inc., San Francisco, California, 1971. *(Cited on page 23.)*

[6] S. Athuraliya, S. H. Low, V. H. Li, and Q. Yin. REM: Active queue management. *IEEE Network*, 15(3):48–53, 2001. *(Cited on page 92.)*

[7] R. Baldick, R. Grant, and E. P. Kahn. Theory and application of linear supply function equilibrium in electricity markets. *Journal of Regulatory Economics*, 2004. To appear. *(Cited on page 166.)*

[8] R. Baldick and W. Hogan. Capacity constrained supply function equilibrium models of electricity markets: stability, non-decreasing constraints, and function space iterations. Working paper PWP-089, University of California Energy Institute, 2001. *(Cited on page 166.)*

[9] R. Baldick and W. Hogan. Stability of supply function equilibria: implications for daily vs. hourly bids in a poolco market. In *POWER 7th Annual Research Conference on Electricity Industry Restructuring*, 2002. *(Cited on page 166.)*

[10] M. Beckmann, C. B. McGuire, and C. B. Winsten. *Studies in the Economics of Transportation*. Yale University Press, New Haven, Connecticut, 1956. *(Cited on pages 155 and 160.)*

[11] J. Bentham. *An Introduction to the Principles and Morals of Legislation*. Clarendon Press, Oxford, United Kingdom, 1907. Originally published in 1823. *(Cited on page 19.)*

[12] J. Bertrand. Théorie des richesses: revue de *Théories mathématiques de la richesse sociale* par Léon Walras et *Recherches sur les principes mathématiques de la théorie des richesses* par Augustin Cournot. *Journal des Savants*, 67:499–508, 1883. *(Cited on page 164.)*

[13] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Cambridge, Massachusetts, second edition, 1999. *(Cited on pages 8, 41, 172 and 227.)*

[14] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, Cambridge, Massachusetts, 2003. *(Cited on pages 7 and 8.)*

[15] D. Bertsimas and J. N. Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, Cambridge, Massachusetts, 1997. *(Cited on pages 68 and 127.)*

[16] F. Bolle. Supply function equilibria and the danger of tacit collusion: the case of spot markets for electricity. *Energy Economics*, 14(2):94–102, 1992. *(Cited on pages 165 and 166.)*

[17] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, United Kingdom, 2003. *(Cited on page 8.)*

[18] B. Briscoe, V. Darlagiannis, O. Heckman, H. Oliver, V. Siris, D. Songhurst, and B. Stiller. A market managed multiservice Internet (M3I). *Computer Communications*, 26(4):404–414, 2003. *(Cited on page 33.)*

[19] C. K. Chau and K. M. Sim. The price of anarchy for non-atomic congestion games with symmetric cost maps and elastic demands. *Operations Research Letters*, 31:327–334, 2003. *(Cited on page 160.)*

[20] E. H. Clarke. Multipart pricing of public goods. *Public Choice*, 11:19–33, 1971. *(Cited on pages 34 and 234.)*

[21] J. R. Correa, A. S. Schulz, and N. Stier Moses. Selfish routing in capacitated networks. *Mathematics of Operations Research*, 2004. To appear. *(Cited on pages 25 and 47.)*

[22] C. A. Courcoubetis, G. D. Stamoulis, C. Manolakis, and F. P. Kelly. An intelligent agent for optimizing QoS-for-money in priced ABR connections. Technical Report 218, The Institute of Computer Science of the Foundation for Research and Technology – Hellas, 1999. *(Cited on page 35.)*

[23] A. A. Cournot. *Recherches sur les Principes Mathématiques de la Théorie des Richesses*. Macmillan and Co., New York, New York, 1897. Translated by N.T. Bacon; originally published in 1838. *(Cited on pages 31, 92 and 164.)*

[24] A. Czumaj and B. Voecking. Tight bounds for worst-case equilibria. In *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 413–420, 2002. *(Cited on page 25.)*

[25] R. E. Dansby and R. D. Willig. Industry performance gradient indexes. *The American Economic Review*, 69(3):249–260, 1979. *(Cited on page 25.)*

[26] P. Dasgupta, P. Hammond, and E. Maskin. The implementation of social choice rules: some general results on incentive compatibility. *Review of Economic Studies*, 46(2):185–216, 1979. *(Cited on page 234.)*

[27] C. d'Aspremont and L.-A. Gérard-Varet. Incentives and incomplete information. *Journal of Public Economics*, 11(1):25–45, 1979. *(Cited on page 234.)*

[28] C. d'Aspremont and L.-A. Gérard-Varet. On Bayesian incentive compatible mechanisms. In J.-J. Laffont, editor, *Aggregation and Revelation of Preferences*, pages 269–288. North-Holland Publishing Company, Amsterdam, The Netherlands, 1979. *(Cited on page 234.)*

[29] A. F. Daughety, editor. *Cournot Oligopoly*. Cambridge University Press, Cambridge, United Kingdom, 1988. *(Cited on page 92.)*

[30] C. J. Day, B. F. Hobbs, and J.-S. Pang. Oligopolistic competition in power networks: a conjectured supply function approach. *IEEE Transactions on Power Systems*, 17(3):597–607, 2002. *(Cited on pages 165 and 166.)*

[31] G. Debreu. *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*. John Wiley and Sons, Inc., New York, New York, 1959. *(Cited on page 23.)*

[32] N. R. Devanur, N. Garg, R. Khandekar, V. Pandit, A. Saberi, and V. V. Vazirani. Price of anarchy, locality gap, and a network service provider game. 2004. In preparation. *(Cited on page 25.)*

[33] P. Dubey. Inefficiency of Nash equilibria. *Mathematics of Operations Research*, 11:1–8, 1986. *(Cited on page 24.)*

[34] B. Duncan. Pumpkin pies and public goods: the raffle fundraising strategy. *Public Choice*, 111(1-2):49–71, 2002. *(Cited on page 35.)*

[35] J. Dupuit. De la mesure de l'utilité des travaux publics. In K. J. Arrow and T. Scitovsky, editors, *Readings in Welfare Economics*, pages 255–283. George Allen and Unwin, Ltd., London, United Kingdom, 1979. Translated by R.H. Barback; originally published in 1844. *(Cited on page 18.)*

[36] R. Edell and P. Varaiya. Providing Internet access: what we learn from INDEX. *IEEE Network*, 13(5):18–25, 1999. *(Cited on page 33.)*

[37] A. Fabrikant, A. Luthra, E. Maneva, C. Papadimitriou, and S. Shenker. On a network creation game. In *Proceedings of the 22nd Annual ACM Symposium on Principles of Distributed Computing*, pages 347–351, 2003. *(Cited on page 25.)*

[38] M. Falkner, M. Devetsikiotis, and I. Lambadaris. An overview of pricing concepts for broadband IP networks. *IEEE Communications Surveys*, 3(2), 2000. *(Cited on page 33.)*

[39] J. Feigenbaum and S. Shenker. Distributed algorithmic mechanism design: recent results and future directions. In *Proceedings of the 6th International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications*, pages 1–13, 2002. *(Cited on page 236.)*

[40] E. J. Friedman and S. Shenker. Learning and implementation on the internet. Preprint, 1998. *(Cited on page 233.)*

[41] J. W. Friedman. *Oligopoly and the Theory of Games*. North-Holland Publishing Company, Amsterdam, The Netherlands, 1977. *(Cited on page 92.)*

[42] D. Fudenberg and D. K. Levine. *The Theory of Learning in Games*. MIT Press, Cambridge, Massachusetts, 1998. *(Cited on page 233.)*

[43] D. Fudenberg and J. Tirole. *Game Theory*. MIT Press, Cambridge, Massachusetts, 1991. *(Cited on pages 8 and 233.)*

[44] R. J. Gibbens and F. P. Kelly. Distributed connection acceptance control for a connectionless network. In *Teletraffic Engineering in a Competitive World: Proceedings of the International Teletraffic Congress (ITC-16)*, pages 941–952, 1999. *(Cited on page 89.)*

[45] R. J. Gibbens and F. P. Kelly. Resource pricing and the evolution of congestion control. *Automatica*, 35:1969–1985, 1999. *(Cited on pages 47 and 92.)*

[46] J. R. Green and J.-J. Laffont. *Incentives in Public Decision-Making*. North-Holland Publishing Company, Amsterdam, The Netherlands, 1979. *(Cited on pages 17, 167 and 234.)*

[47] R. J. Green. Increasing competition in the British electricity spot market. *Journal of Industrial Economics*, 44(2):205–216, 1996. *(Cited on page 165.)*

[48] R. J. Green and D. M. Newbery. Competition in the British electricity spot market. *Journal of Political Economy*, 100(5):929–953, 1992. *(Cited on pages 165 and 166.)*

[49] S. J. Grossman. Nash equilibrium and the industrial organization of markets with large fixed costs. *Econometrica*, 49(5):1149–1172, 1981. *(Cited on page 165.)*

[50] T. Groves. Incentives in teams. *Econometrica*, 41(4):617–631, 1973. *(Cited on pages 34 and 234.)*

[51] A. Gupta, D. O. Stahl, and A. B. Whinston. The economics of network management. *Communications of the ACM*, 42(9):57–63, 1999. *(Cited on page 34.)*

[52] B. Hajek and G. Gopalakrishnan. Do greedy autonomous systems make for a sensible Internet? Presented at the Conference on Stochastic Networks, Stanford University, 2002. *(Cited on pages 35, 40 and 43.)*

[53] B. Hajek and S. Yang. Strategic buyers in a sum-bid game for flat networks. Submitted, 2004. *(Cited on pages 84 and 135.)*

[54] O. D. Hart. Imperfect competition in general equilibrium: An overview of recent work. In K. J. Arrow and S. Honkapohja, editors, *Frontiers of Economics*, pages 100–149. Basil Blackwell, Oxford, United Kingdom, 1985. *(Cited on page 165.)*

[55] A. Haurie and P. Marcotte. On the relationship between Nash-Cournot and Wardrop equilibria. *Networks*, 15(1):295–308, 1985. *(Cited on page 155.)*

[56] B. F. Hobbs, M. H. Rothkopf, L. C. Hyde, and R. P. O'Neill. Evaluation of a truthful revelation auction in the context of energy markets with nonconcave benefits. *Journal of Regulatory Economics*, 18(1):5–32, 2000. *(Cited on page 236.)*

[57] M. O. Jackson and J. M. Swinkels. Existence of equilibrium in auctions and discontinuous Bayesian games: endogenous and incentive compatibility sharing rules. Social Science Working Paper 1075, Division of the Humanities and Social Sciences, California Institute of Technology, 1999. *(Cited on page 64.)*

[58] R. Johari, S. Mannor, and J. N. Tsitsiklis. Efficiency loss in a network resource allocation game: the case of elastic supply. 2004. In preparation. *(Cited on page 8.)*

[59] R. Johari and D. K. H. Tan. End-to-end congestion control for the Internet: delays and stability. *IEEE/ACM Transactions on Networking*, 9(6):818–832, 2001. *(Cited on page 91.)*

[60] R. Johari and J. N. Tsitsiklis. Efficiency loss in a network resource allocation game. *Mathematics of Operations Research*, 2004. To appear. Portions reprinted with permission, © 2004, The Institute for Operations Research and the Management Sciences, 901 Elkridge Landing Road, Suite 400, Linthicum, Maryland, 21090-2909, USA. *(Cited on page 8.)*

[61] F. P. Kelly. Network routing. *Philosophical Transactions: Physical Sciences and Engineering*, 337(1647):343–367, 1991. *(Cited on page 155.)*

[62] F. P. Kelly. Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, 8:33–37, 1997. *(Cited on pages 22, 26, 31, 34, 35, 36, 37, 38, 40, 62, 64, 81, 82, 83, 84, 89, 91, 92 and 95.)*

[63] F. P. Kelly. Models for a self managed Internet. *Philosophical Transactions: Mathematical, Physical, and Engineering Sciences*, 358(1773):2335–2348, 2000. *(Cited on pages 21 and 92.)*

[64] F. P. Kelly. Stability and fairness of end-to-end congestion control. *European Journal of Control*, 9:159–176, 2003. *(Cited on page 91.)*

[65] F. P. Kelly, A. K. Maulloo, and D. K. Tan. Rate control for communication networks: shadow prices, proportional fairness, and stability. *Journal of the Operational Research Society*, 49:237–252, 1998. *(Cited on pages 22, 31, 91, 92, 93, 97, 98, 135 and 232.)*

[66] P. B. Key. Service differentiation: Congestion pricing, brokers and bandwidth futures. Presented at Network and Operating System Support for Digital Audio and Video (NOSSDAV), 1999. *(Cited on page 89.)*

[67] P. B. Key and D. R. McAuley. Differential QoS and pricing in networks: where flow control meets game theory. *IEE Proceedings – Software*, 146(1), 1999. *(Cited on page 233.)*

[68] P. D. Klemperer and M. A. Meyer. Price competition vs. quantity competition: the role of uncertainty. *The RAND Journal of Economics*, 17(4):618–638, 1986. *(Cited on page 165.)*

[69] P. D. Klemperer and M. A. Meyer. Supply function equilibria in oligopoly under uncertainty. *Econometrica*, 57(6):1243–1277, 1989. *(Cited on pages 21, 40, 165 and 166.)*

[70] E. Koutsoupias and C. Papadimitriou. Worst-case equilibria. In *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science*, pages 404–413, 1999. *(Cited on page 25.)*

[71] S. Kunniyur and R. Srikant. Analysis and design of an adaptive virtual queue (AVQ) algorithm for active queue management. In *Proceedings of ACM SIGCOMM*, pages 123–134, 2001. *(Cited on page 92.)*

[72] S. Kunniyur and R. Srikant. End-to-end congestion controls schemes: utility functions, random losses, and ECN marks. *IEEE/ACM Transactions on Networking*, 11(5):689–702, 2003. *(Cited on page 92.)*

[73] R. J. La and V. Anantharam. Charge-sensitive TCP and rate control in the Internet. In *Proceedings of IEEE INFOCOM*, pages 1166–1175, 2000. *(Cited on pages 35 and 47.)*

[74] R. J. La and V. Anantharam. Utility based rate control in the Internet for elastic traffic. *IEEE/ACM Transactions on Networking*, 10(2):272–286, 2002. *(Cited on page 35.)*

[75] J.-J. Laffont and E. Maskin. The theory of incentives: an overview. In *Advances in Economic Theory: 4th World Congress of the Econometric Society*, pages 31–94, 1982. *(Cited on page 234.)*

[76] A. A. Lazar and N. Semret. Design and analysis of the Progressive Second Price auction for network bandwidth sharing. Technical report 487-98-21, Columbia University Center for Telecommunications Research, 1999. *(Cited on pages 34 and 235.)*

[77] J. K. MacKie-Mason and H. R. Varian. Pricing congestible network resources. *IEEE Journal on Selected Areas in Communications*, 13(7):1141–1149, 1995. *(Cited on page 33.)*

[78] J. K. MacKie-Mason and H. R. Varian. Pricing the Internet. In B. Kahin and J. Keller, editors, *Public Access to the Internet*, pages 269–314. MIT Press, Cambridge, Massachusetts, 1995. *(Cited on page 34.)*

[79] R. T. Maheswaran and T. Basar. Nash equilibrium and decentralized negotiation in auctioning divisible resources. *Group Decision and Negotiation*, 12(5):361–395, 2003. *(Cited on pages 47 and 68.)*

[80] A. Manning. *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton University Press, Princeton, New Jersey, 2003. *(Cited on page 92.)*

[81] A. Marshall. *Principles of Economics*. Macmillan and Co., London, United Kingdom, 1920. *(Cited on pages 18, 19 and 23.)*

[82] A. Mas-Colell, M. D. Whinston, and J. R. Green. *Microeconomic Theory*. Oxford University Press, Oxford, United Kingdom, 1995. *(Cited on pages 8, 17, 18, 22, 23, 36, 40, 47, 57, 135, 148, 159, 167, 170, 181, 232 and 233.)*

[83] E. Maskin. Nash equilibrium and welfare optimality. *Review of Economic Studies*, 66(1):23–38, 1999. *(Cited on page 234.)*

[84] M. Mavronicolas and P. Spirakis. The price of selfish routing. In *Proceedings of the 33rd Annual ACM Symposium on the Theory of Computing*, pages 510–519, 2001. *(Cited on page 25.)*

[85] J.-F. Mertens, S. Sorin, and S. Zamir. Repeated games. Discussion Papers 9420-9422, Université Catholique de Louvain Center for Operations Research and Econometrics, 1994. *(Cited on page 233.)*

[86] D. Monderer and L. S. Shapley. Potential games. *Games and Economic Behavior*, 14(1):124–143, 1996. *(Cited on page 47.)*

[87] J. Moore. Implementation, contracts, and renegotiation in environments with complete information. In *Advances in Economic Theory: 6th World Congress of the Econometric Society*, pages 182–282, 1992. *(Cited on page 234.)*

[88] R. B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6:58–73, 1981. *(Cited on pages 59 and 234.)*

[89] R. B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, Cambridge, Massachusetts, 1991. *(Cited on page 8.)*

[90] J. F. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences of the United States of America*, 36(1):48–49, 1950. *(Cited on page 24.)*

[91] New England Demand Response Initiative. Dimensions of demand response: report and recommendations of the New England Demand Response Initiative. 2003. *(Cited on page 164.)*

[92] W. Novshek. On the existence of Cournot equilibrium. *Review of Economic Studies*, 52(1):85–98, 1985. *(Cited on page 139.)*

[93] A. M. Odlyzko. Paris Metro Pricing for the Internet. In *Proceedings of the ACM Conference on Electronic Commerce (EC '99)*, pages 140–147, 1999. *(Cited on page 33.)*

[94] A. M. Odlyzko. Internet pricing and the history of communications. *Computer Networks*, 36(5-6):493–517, 2001. *(Cited on page 33.)*

[95] A. M. Odlyzko. Pricing and architecture of the Internet: historical perspectives from telecommunications and transportation. 2004. Preprint. *(Cited on page 90.)*

[96] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, Cambridge, Massachusetts, 1994. *(Cited on pages 8, 24 and 210.)*

[97] A. Ozdaglar and D. Acemoglu. Flow control, routing, and performance from service provider viewpoint. 2004. Submitted. *(Cited on page 59.)*

[98] T. R. Palfrey. Implementation in Bayesian equilibrium: the multiple equilibrium problem in mechanism design. In *Advances in Economic Theory: 6th World Congress of the Econometric Society*, pages 283–323, 1992. *(Cited on page 234.)*

[99] C. Papadimitriou. Algorithms, games, and the Internet. In *Proceedings of the 33rd Annual ACM Symposium on the Theory of Computing*, pages 749–753, 2001. *(Cited on pages 25 and 195.)*

[100] G. Perakis. The "price of anarchy" under nonlinear and asymmetric costs. In preparation, 2004. *(Cited on page 25.)*

[101] A. C. Pigou. *The Economics of Welfare*. Macmillan and Co., London, United Kingdom, 1932. Originally published in 1920. *(Cited on page 24.)*

[102] K. Ramakrishnan, S. Floyd, and D. Black. The addition of explicit congestion notification (ECN) to IP. Request for Comments 3168, Internet Engineering Task Force, 2001. *(Cited on page 92.)*

[103] R. T. Rockafellar. *Convex Analysis*. Priceton Landmarks in Mathematics and Physics. Princeton University Press, Princeton, New Jersey, 1970. *(Cited on pages 7, 8, 77, 103 and 131.)*

[104] J. Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*, 33(3):520–534, 1965. *(Cited on pages 69, 103, 139 and 193.)*

[105] R. W. Rosenthal. A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory*, 2:65–67, 1973. *(Cited on page 47.)*

[106] T. Roughgarden. The price of anarchy is independent of the network topology. In *Proceedings of the 34th Annual ACM Symposium on the Theory of Computing*, pages 428–437, 2002. *(Cited on pages 25 and 81.)*

[107] T. Roughgarden. *Selfish Routing*. PhD thesis, Cornell University, 2002. *(Cited on pages 25 and 94.)*

[108] T. Roughgarden and E. Tardos. How bad is selfish routing? *Journal of the ACM*, 49(2):236–259, 2002. *(Cited on pages 25, 47, 55, 94, 155 and 160.)*

[109] T. Roughgarden and E. Tardos. Bounding the inefficiency of equilibria in nonatomic congestion games. *Games and Economic Behavior*, 47(2):389–403, 2004. *(Cited on page 25.)*

[110] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, Inc., New York, New York, 1964. *(Cited on page 8.)*

[111] A. Rudkevich, M. Duckworth, and R. Rosen. Modeling electricity pricing in a deregulated generation industry: the potential for oligopoly pricing in a poolco. *Energy Journal*, 19(3):19–48, 1998. *(Cited on page 166.)*

[112] S. Sanghavi and B. Hajek. Optimal allocation of a divisible good to strategic buyers. 2004. Submitted. *(Cited on page 211.)*

[113] A. S. Schulz and N. Stier Moses. On the performance of user equilibria in traffic networks. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 86–87, 2003. *(Cited on page 25.)*

[114] A. S. Schulz and N. Stier Moses. The price of anarchy with elastic demands. In preparation, 2004. *(Cited on pages 25 and 160.)*

[115] F. C. Schweppe, M. C. Caramanis, R. D. Tabors, and R. E. Bohn. *Spot Pricing of Electricity*. Kluwer Academic Publishers, Boston, Massachusetts, 1988. *(Cited on page 186.)*

[116] N. Semret. *Market Mechanisms for Network Resource Sharing*. PhD thesis, Columbia University, 1999. *(Cited on pages 34 and 235.)*

[117] A. K. Sen. *Collective Choice and Social Welfare*. North-Holland Publishing Company, Amsterdam, The Netherlands, 1979. *(Cited on page 19.)*

[118] A. K. Sen. *Rationality and Freedom*. The Belknap Press of Harvard University Press, Cambridge, Massachusetts, 2002. *(Cited on page 19.)*

[119] C. Shapiro. Theories of oligopoly behavior. In R. Schmalensee and R. D. Willig, editors, *Handbook of Industrial Organization*, volume 1, pages 329–414. Elsevier Science, Amsterdam, The Netherlands, 1989. *(Cited on pages 25, 92 and 164.)*

[120] Y. Sheffi. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, Englewood Cliffs, New Jersey, 1985. *(Cited on page 160.)*

[121] S. Shenker. Fundamental design issues for the future Internet. *IEEE Journal on Selected Areas in Communications*, 13:1176–1188, 1995. *(Cited on page 37.)*

[122] S. Shenker, D. Clark, D. Estrin, and S. Herzog. Pricing in computer networks: reshaping the research agenda. *Telecommunications Policy*, 20(3):183–201, 1996. *(Cited on page 33.)*

[123] J. Shu and P. Varaiya. Pricing network services. In *Proceedings of IEEE INFOCOM*, pages 1221–1230, 2003. *(Cited on page 34.)*

[124] D. Smart. *Fixed Point Theorems*. Cambridge University Press, Cambridge, United Kingdom, 1974. *(Cited on page 159.)*

[125] A. Smith. *An Inquiry Into the Nature and Causes of the Wealth of Nations*. The Modern Library, New York, New York, 1937. Originally published in 1776. *(Cited on page 15.)*

[126] D. J. Songhurst, editor. *Charging Communication Networks: From Theory to Practice*. Elsevier Science, Amsterdam, The Netherlands, 1999. *(Cited on page 33.)*

[127] R. Srikant. *The Mathematics of Internet Congestion Control*. Birkhauser, Boston, Massachusetts, 2004. *(Cited on pages 91 and 232.)*

[128] G. J. Stigler. The development of utility theory I. *Journal of Political Economy*, 58(4):307–327, 1950. *(Cited on page 19.)*

[129] G. J. Stigler. The development of utility theory II. *Journal of Political Economy*, 58(5):373–396, 1950. *(Cited on page 19.)*

[130] G. J. Stigler. *The Theory of Price*. Macmillan and Co., New York, New York, 1987. *(Cited on page 23.)*

[131] S. Stoft. *Power System Economics: Designing Markets for Electricity*. IEEE Press, Piscataway, New Jersey, 2002. *(Cited on pages 20 and 164.)*

[132] I. Stoica, H. Abdel-Wahab, K. Jeffay, S. Baruah, J. Gehrke, and C. Plaxton. A proportional share resource allocation algorithm for real-time time-shared systems. In *Proceedings of the 17th IEEE Real-Time Systems Symposium*, pages 288–299, 1996. *(Cited on page 35.)*

[133] C. Supatgiat, R. Q. Zhang, and J. R. Birge. Equilibrium values in a competitive power exchange market. *Computational Economics*, 17(1):93–121, 2001. *(Cited on page 166.)*

[134] J. Tirole. *The Theory of Industrial Organization*. MIT Press, Cambridge, Massachusetts, 1988. *(Cited on pages 18, 24 and 135.)*

[135] United States Department of Justice and the Federal Trade Commission. Horizontal merger guidelines. 1992. *(Cited on page 24.)*

[136] United States Federal Energy Regulatory Commission. Notice of proposed rulemaking. Docket No. RM01-12-000, 2002. *(Cited on page 165.)*

[137] H. R. Varian. *Microeconomic Analysis*. W.W. Norton and Co., New York, New York, 1992. *(Cited on pages 8, 23, 106, 207 and 232.)*

[138] A. Vetta. Nash equilibrium in competitive societies, with applications to facility location, traffic routing, and auctions. In *Proceedings of the 43rd Annual IEEE Symposium on the Foundations of Computer Science*, pages 416–425, 2002. *(Cited on page 25.)*

[139] W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance*, 16(1):8–37, 1961. *(Cited on pages 34, 49 and 234.)*

[140] G. Vinnicombe. On the stability of networks operating TCP-like congestion control. In *Proceedings of IFAC World Congress*, 2002. *(Cited on page 91.)*

[141] N.-H. M. von der Fehr and D. Harbord. Spot market competition in the U.K. electricity industry. *Economic Journal*, 103(418):531–546, 1993. *(Cited on page 166.)*

[142] L. Walras. *Éléments d'Économie Politique Pure; ou, Théorie de la Richesse Sociale*. Richard D. Irwin, Inc., Homewood, Illinois, 1954. Translated by W. Jaffé; originally published in 1874. *(Cited on page 23.)*

[143] J. J. Wang and J. F. Zender. Auctioning divisible goods. *Economic Theory*, 19(4):673–705, 2002. *(Cited on page 39.)*

[144] J. G. Wardrop. Some theoretical aspects of road traffic research. In *Proceedings of the Institute of Civil Engineers, Part II*, volume 1, pages 325–378, 1952. *(Cited on pages 94, 155 and 160.)*

[145] P. Whittle. *Optimization Under Constraints: Theory and Applications of Nonlinear Programming*. John Wiley and Sons, Inc., New York, New York, 1971. *(Cited on page 8.)*

[146] R. Wilson. Auctions of shares. *Quarterly Journal of Economics*, 93(4):675–689, 1979. *(Cited on pages 21, 39 and 165.)*

[147] R. Wilson. Architecture of power markets. *Econometrica*, 70(4):1299–1340, 2002. *(Cited on pages 166 and 236.)*

[148] S. Yang and B. Hajek. An efficient mechanism for allocation of a divisible good and its application to network resource allocation. 2004. In preparation. *(Cited on page 211.)*