Essays on Variational Inequalities and Competitive Supply Chain Models

by

Marina Zaretsky

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Operations Research

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2004

Author ....
Sloan School of Management
August 2, 2004

Certified by ....
Georgia Perakis
Sloan Career Development Associate Professor of Operations Research
Thesis Supervisor

Accepted by ....
James B. Orlin
Edward Pennell Brooks Professor of Operations Research
Codirector of Operations Research Center

Essays on Variational Inequalities and Competitive Supply Chain Models

by

Marina Zaretsky

Submitted to the Sloan School of Management
on August 2, 2004, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Operations Research

Abstract

In the first part of the thesis we combine ideas from cutting plane and interior point methods to solve variational inequality problems efficiently. In particular, we introduce "smarter" cuts into two general methods for solving these problems. These cuts utilize second order information on the problem through the use of a gap function. We establish convergence results for both methods, as well as complexity results for one of the methods. Finally, we compare the performance of an approach that combines affine scaling and cutting plane methods with other methods for solving variational inequalities.

The second part of the thesis considers a supply chain setting where several capacitated suppliers compete for orders from a single retailer in a multi-period environment. At each period the retailer places orders to the suppliers in response to the prices and capacities they announce. Our model allows the retailer to carry inventory. Furthermore, suppliers can expand their capacity at an additional cost; the retailer faces exogenous, price-dependent, stochastic demand. We analyze discrete as well as continuous time versions of the model: (i) we illustrate the existence of equilibrium policies; (ii) we characterize the structure of these policies; (iii) we consider coordination mechanisms; and (iv) we present some computational results. We also consider a modified model that uses option contracts and finally present some extensions.

Thesis Supervisor: Georgia Perakis
Title: Sloan Career Development Associate Professor of Operations Research

Acknowledgments

This thesis would not be possible without the continuous encouragement and support from my family and my advisor. My family had patience to accept my pursuing of the degree and were always a resource to rely on even in the most unexpected situations. My advisor generously shared her time and expertise, and I learned a lot from her. But not only about research per se, she also patiently endured my vacillations and taught me about the importance of the positive attitude towards research. Her energy and enthusiasm have been always stimulating.

Aside from providing me with the research experience, the years at MIT were a process of learning more about myself as a human in this world, and where, if not in the academia with its mix of bureaucracy, ambition, brilliance, and research, is one provided with one the best environments to do this? Independent of whether my experiences were positive or negative, they still added to my growing. On the positive note, I would like to mention the MIT ballroom team, that provided a great environment to wind down after stressful days and also allowed me to improve my dance skills and interact with a coherent and dedicated group of dancers. I am in particular indebted to my last partner, working with whom reaffirmed that trust, enthusiasm, and perseverance can create success if not magic.

Finally, I especially would like to thank my boyfriend for helping me survive through MIT experiences. He was there for me (though most of the time on the other end of a telephone line) when I needed emotional support no matter how serious a crisis. I am lucky to have you bringing precious kindness and joy into my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Variational inequality problems ($VIPs$) arise frequently in a variety of application areas ranging from transportation and telecommunication to finance and economics. Moreover, variational inequalities provide a unifying framework for studying a number of important mathematical programming problems including equilibrium, minimax, saddle point, complementarity and optimization problems. As a result, variational inequalities have been the subject of extensive research over the past forty years. In particular, a variational inequality problem seeks a point

$$x^* \in K \text{ such that } f(x^*)'(x - x^*) \geq 0, \text{ for all } x \in K, \tag{1.1}$$

where $K \subseteq \mathbb{R}^n$ is the ground set and $f : K \to \mathbb{R}^n$ is the problem function. That is variational inequality problem seeks a point, at which function value forms acute angle with every feasible direction from that point as shown in Figure 1-1 below.

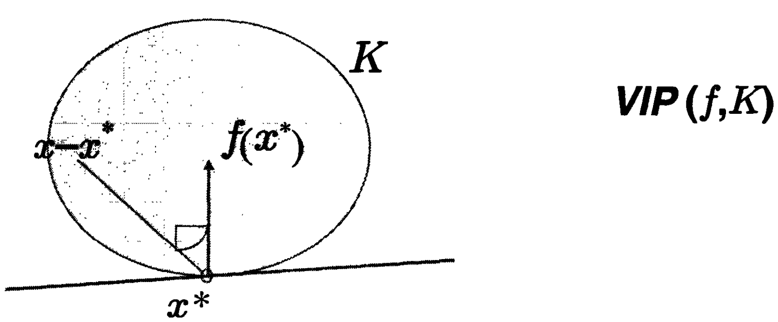

Figure 1-1: Variational inequality problem

Many problems arising in the decentralized setting, i.e., where many agents compete can be cast as generalized Nash equilibrium problems (GNEPs). Examples include competitive supply chain management. In particular, a GNEP seeks to find an optimal policy p_i for a player i, so that each player's payoff function is maximized at p_i given that other players keep their policies at an equilibrium level, i.e.

$$p_i = \arg \max_{s_i \in K_i(p_{-i})} \pi(s_i, p_{-i}), \tag{1.2}$$

where $\pi(s_i, p_{-i})$ is ith player's payoff function and $K_i(p_{-i})$ is a feasible set of the ith player's strategies; s_i denotes the i'the player strategy, and p_{-i} denotes the equilibrium strategies of all the players other than i.

As the overall welfare, that a system with competing, payoff maximizing agents achieves, is often worse than the welfare of the system when it would act as if managed by a single agent, the researchers face the questions:

- Can we induce individual agent to replicate the policies of a single decision-maker?
- Can we suggest contract among agent in supply chain such that the decentralized (Nash) equilibrium solution is the same as the one by a single optimizing decision-maker?

The first part of this thesis studies the efficient solution of VIPs, while the second part studies equilibrium problems as they arise in supply chain management.

In this thesis we study equilibrium problems as they arise in competitive supply chain. We reformulate the problems as VIPs (or more precisely quasi-variational inequality problems ($QVIP$s)). Therefore, general results and algorithms for VIPs are applicable to these models. The second part of the work can be viewed as both an application of VIPs in economics, in particular, in industrial organization, and as a concrete model of competitive interactions in supply chain. Modelling competition has recently emerged as a popular topic of research in management science (see [3], [6], [7], [9], [31] for example.) VIPs in the context of supply chain also have been studied by many researchers. A review of VIP models for perfect competition in oligopolies in static setting can be found in [46]. The models in [46] assume that the system clears the market, i.e., demand equals supply at each agent. In this thesis we take a different approach, we assume that every agent is a profit maximizer, while the system faces exogenous, price dependent demand. As a result, some of the demand might not be satisfied. Pang [50], Pang and Fukushima [49] have recently studied VIP models for oligopolies in a static setting. These papers also provide further references to how VIP models can be used to model equilibrium problems. Unlike these authors, instead of focusing on how the methods for solving VIPs can be extended to solve for equilibria in competitive supply chain problems, we concentrate on characterizing the equilibria in a dynamic (multi-period) oligopoly setting. We also study both continuous and discrete dynamics. We employ ideas from dynamic programming, game theory and differential game theory (see [4], [29] [24], and [20] for more references).

Both of the parts of this thesis are united by the topic of variational inequalities. The first part contributes to the development of algorithms for solving VIPs and the second part contributes to studying a particular model of supply chain competition, which can be analyzed using VIP theory and algorithms. Moreover, the second part of the thesis contributes to the literature on developing contracts in a competitive supply chain setting (in particular, we study equilibrium policies, coordination and the dynamics of the competition). and the dynamics of the competition).

Chapter 2

Variational Inequalities: Complexity and Convergence Results in Methods with Cuts

2.1 Introduction

Our goal in this chapter is to achieve better theoretical and computational efficiency by introducing "smarter" cuts whose definition is based on the dual gap function associated with a VIP. We introduce and study these cuts in the context of a cutting plane method (the general geometric framework as in Magnanti and Perakis [39]) and an interior point method (the affine scaling method as in Gonzaga and Carlos [28]).

Cutting plane methods have been used extensively in the literature for solving optimization problems. As applied to variational inequality problems, these methods include among others the ellipsoid method by Lüthi [37], the general geometric framework by Magnanti and Perakis [39], analytic center methods (see, for example, Goffin et. al. [26]), and barrier methods (see, for example, Nesterov and Nemirovskiy [47], and Nesterov and Vial [48]). Cutting plane methods incorporate several types of cuts: linear cuts (see, for example, Goffin et. al. [26]), quadratic cuts (see, for example, Lüthi and Büeler [38], Denault and Goffin [18]) and nonlinear cuts (see, for example, Nesterov and Nemirovskiy [48]). These methods converge to a solution when the problem function satisfies some form of a monotonicity condition. Moreover, for some of these methods, researchers have established complexity results (see, for example, [39], [48]). In this thesis, we will consider an extension of the general geometric framework [39] (GGF) that can also incorporates nonlinear cuts. The motivation comes from the fact that this framework encapsulates several well-known methods for solving optimization problems such as the ellipsoid method, the volumetric center method, and the method of centers of gravity. Furthermore, an additional motivation comes from the fact that complexity bounds have been established for this framework.

Although one can establish complexity bounds for cutting plane methods, these methods are often computationally expensive in practice. This is due to the fact that the complexity bounds established are often tight in practice but also due to the fact that most cutting plane methods require the computation of a "nice" set and its "center" at each iteration. As a result, we also consider alternate methods for solving variational inequalities such as interior point methods. The motivation in considering this class of methods comes from the observation that the methods have been successful in solving linear optimization problems in practice. For several methods in this class, researchers have established complexity bounds. Consequently, these methods might be attractive for solving other problem classes as well. In particular, in this thesis we consider versions of the affine scaling method (AS). This method was originally developed for solving linear optimization problems by Dikin in 1967 [19]. Subsequently, Ye [61], [62] and more recently Tseng [57] extended this method for solving quadratic optimization problems, and Sun [54] and Gonzaga and Carlos [28], for solving convex optimization problems. Our motivation in studying this method, as it applies to variational inequalities, comes from its simplicity. Moreover, the version of the method that we study in this thesis is motivated by the Frank-Wolfe method [22], which is widely used by transportation practitioners.

We believe this chapter contributes to the existing body of literature by

- introducing cuts that take advantage of second order information in the VIP function;

- presenting polynomially convergent methods;

- proposing more practical methods that are easy to perform computationally.

This chapter is organized as follows: in the remainder of this section we provide some background and describe some useful concepts. In Section 2.2, we describe the cuts that are constructed based on information from the dual gap function. In Section 2.3, we establish complexity results for the GGF with such cuts. In Section 2.4, we focus on the solution of symmetric, monotone VIPs, using an affine scaling method. We provide some convergence results and propose combining affine scaling with cuts. This allows us to suggest schemes that are more tractable computationally. In Section 2.5, we summarize our conclusions.

2.1.1 Preliminaries

In this section we review some basic definitions. We first define the notion of a weak variational inequality problem and relate it to a variational inequality problem.

Definition 1 Let $f : K \rightarrow \mathbb{R}^n$ be a given n-dimensional function and let $K \subset \mathbb{R}^n$ be a given ground set. A point $\bar{x} \in K$ is a *weak VIP solution* if for all $x \in K$, $f(x)'(x - \bar{x}) \geq 0$. We will refer to the problem that seeks a weak VIP solution as a *weak variational inequality problem* ($WVIP$).

Definition 2

1. A function f is *quasimonotone* on K if $f(y)'(x - y) > 0$ implies that $f(x)'(x - y) \geq 0$, for all $x, y \in K$.

2. A function f is *pseudomonotone* on K if $f(y)'(x - y) \geq 0$ implies that $f(x)'(x - y) \geq 0$, for all $x, y \in K$.

3. A function f is *monotone* if $(f(x) - f(y))'(x - y) \geq 0$, for all $x, y \in K$.

Assumption 1 *K is a closed, bounded and convex set with a nonempty interior.*

In particular, through Assumption 1 we assume that there exist positive constants L and l such that the feasible region is contained in a ball of radius 2^L and, in turn, contains a ball of radius 2^{-l} (see [39] for a further discussion on how to explicitly define these constants for a polyhedral feasible region).

Lemma 1 *When the problem function f is continuous, a VIP is equivalent to a $WVIP$ if one of the following conditions holds:*

(a) The underlying problem function f is quasimonotone and for some $y \in K$, $f(x^*)'(y - x^*) > 0$;

(b) The underlying problem function f is pseudomonotone.

The proof of this lemma as well as those of other unreferenced results are contained in Perakis and Zaretsky [51]. In the rest of this chapter we make the following assumptions.

Assumption 2 *Problem function f is continuous on K.*

2.2 Cuts based on gap functions

In this section we introduce cuts based on the dual gap function associated with the VIP. We first notice that the VIP and $WVIP$ solutions can be characterized through appropriate gap functions.

Definition 3 We define function $C_p(y) = \max_{z \in K} f(y)'(y - z)$ as the *primal gap function* and function $C_d(y) = \max_{z \in K} f(z)'(y - z)$ as the *dual gap function.*

As is well known, the VIP solution set X^* consists of points $\arg\min_{x \in K} C_p(x)$, i.e., $X^* = \{x \mid C_p(x) = 0\}$. Moreover, the gap function C_d is a closed convex function, that is strictly positive outside the solution set. The solution set of a $WVIP$, coincides with the set $\arg\min_{x \in K} C_d(x) = \{x \mid C_d(x) = 0\}$. Finally, we note that when the conditions of Lemma 1 are satisfied, then the VIP solution set coincides with the $WVIP$ solution set.

In Sections 3 and 4, we aim to utilize the connection between the dual gap and VIP functions, to construct cuts and employ them in two schemes for solving VIPs. We note that such cuts will exploit second order information, unlike for example linear cuts of the type $Cut(x) = \{y \in K \mid f(x)'(y-x) \leq 0\}$ that use only the slope $f(x)$ to define a cut through a point $x \in K$ (these linear cuts are often considered in the literature, see [38] or [39]).

2.2.1 Linear cuts via an exact dual gap function

To motivate our results, we first consider the following type of cuts. Let

$$y_x = \arg\max_{y \in K} f(y)'(x-y), \tag{2.1}$$

and $Cut(y,x) = \{z \in K \mid f(y)'(z-x) \leq 0\}$ be the half space through point x with slope $f(y)$.

The following lemma asserts that cuts $Cut(y_x, x)$ contain the $WVIP$ solution set.

Lemma 2 $\{z \in K \mid C_d(z) \leq C_d(x)\} \subseteq Cut(y_x, x)$.

Proof. Let y_x, y_z be as defined in (2.1), then when $C_d(z) \leq C_d(x)$, the following holds.

$$\begin{aligned} f(y_x)'(z-x) &= f(y_x)'(z-y_x) + f(y_x)'(y_x-x) \leq f(y_z)'(z-y_z) + f(y_x)'(y_x-x), \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(using (2.1))} \\ &\leq f(y_x)'(x-y_x) + f(y_x)'(y_x-x), \text{ (using } C_d(z) \leq C_d(x)) \\ &= 0. \blacksquare \end{aligned}$$

Determining the slope of the cut $Cut(y_x, x)$ can be computationally expensive. This is due to the nonlinearity and, perhaps, even the non-convexity of the subproblem that generates point y_x. As a result, next we consider schemes that compute approximations of the direction $f(y_x)$, yet generating sequences that converge to a $WVIP$ solution.

2.2.2 Linear cuts via approximations of a dual gap function

Motivation

To motivate the linear cuts we introduce in this section, we first consider an approximation of the gap function C_d. In particular, given a point $y \in K$, the mean value theorem applied to the function $f(\cdot)'(y-x)$ around point $x \in K$ implies that $f(y)'(y-x) = f(x)'(y-x) + (y-x)'\nabla f(z)'(y-x)$, for some $z \in [x;y]$. As a result, we can rewrite subproblem (2.1) as $y_x = \arg\max_{y \in K} f(y)'(x-y) = \arg\max_{y\in K}\left(f(x)'(x-y) - (y-x)'\nabla f(z)(y-x)\right)$ for some $z \in [x;y]$. Motivated by this observation, we will consider cuts whose directions will be determined from this approximation. To develop this observation more formally, we first need to impose an additional assumption on the Jacobian matrix.

Definition 4 A function $f : \mathbb{R}^n \to \mathbb{R}^n$ has the property of *Jacobian similarity* if the Jacobian matrix is positive semi-definite and there exists some constant $\rho \geq 1$ such that the Jacobian matrix satisfies $d'\nabla f(y)d \leq \rho d'\nabla f(x)d$, for all $x, y \in K$, and $d \in \mathbb{R}^n$.

Assumption 3 *Problem function f satisfies the Jacobian similarity property.*

Remarks.

1. In the context of convex optimization, there is an analogous property called *Hessian similarity* (see [54]).

2. The Jacobian similarity property holds for monotone functions with a bounded Jacobian matrix. This property is similar to the property of self-concordance as it applies to barrier functions. In the context of nonlinear optimization, Nemirovskiy and Nesterov [47] have shown that if the self-concordance property holds for a barrier function, then the Hessian similarity property also holds locally.

3. When the problem function f is strictly monotone, one can choose the Jacobian similarity constant ρ as the ratio of the eigenvalues of the symmetrized Jacobian matrix $\frac{\nabla f^t + \nabla f}{2}$, that is, $\rho = \frac{\sup_{y\in K}\lambda_{\max}\left(\nabla f^t(y) + \nabla f(y)\right)}{\inf_{x \in K}\lambda_{\min}(\nabla f^t(x) + \nabla f(x))}$.

4. Finally, notice that when the problem function f is affine, then we can choose $\rho = 1$.

We let $H(x) = \frac{\nabla f(x)+\nabla f(x)'}{2}$ denote the symmetrized Jacobian matrix of f and let $\|w\|^2_{H(x)} = w'H(x)w$. Assumption 3 then implies that $f(y)'(y-x) = f(x)'(y-x) + (y-x)'\nabla f(z)'(y-x) \leq f(x)'(y-x) + \rho \|x-y\|^2_{H(x)}$, where $z \in [x;y]$. Therefore, $f(y)'(x-y) \geq f(x)'(x-y) - \rho \|x-y\|^2_{H(x)}$.

First approximation

The previous discussion motivates us to consider cuts whose directions are determined by solving a quadratic approximation of subproblem (2.1). In particular, we approximate the gap function $C_d(x)$ with its quadratic approximation

$$C_d^1(x) = \max_{y\in K}\left(f(x)'(x-y) - \rho \|x-y\|^2_{H(x)}\right).$$

Let the point y_x^1 be the maximizer of $C_d^1(x)$. We state properties of this type of cuts in the following two propositions.

Lemma 3 *Suppose Assumptions 2 and 3 hold. The cuts $Cut(y_x^1, x)$ contain the WVIP solution set, i.e.* $\{z \in K \mid C_d(z) \leq C_d^1(x)\} \subseteq \{z \in K \mid f(y_x^1)'(z-x) \leq 0\}$.

Under Assumptions 1 - 3, the WVIP solution set X^* coincides with the set of minimizers of $C_d^1(x)$.

Proof. Notice that inequality $C_d(z) \leq C_d^1(x)$ holds if and only if for all $y \in K$,

$$f(y)'(z-y) \leq f(x)'(x-y_x^1) - \rho \left\|x-y_x^1\right\|^2_{H(x)}. \tag{2.2}$$

An application of the mean value theorem together with Assumption 3 (i.e., the Jacobian similarity property) imply that there exists a point $w_x \in [x; y_x^1]$ such that

$$f(x)'(x-y_x^1) - \left\|x-y_x^1\right\|^2_{H(w_x)} = f(y_x^1)'(x-y_x^1). \tag{2.3}$$

Therefore, if z is such that $C_d(z) \leq C_d^1(x)$, then

$$f(y_x^1)'(z-x) = f(y_x^1)'(z-y_x^1) + f(y_x^1)'(y_x^1-x)$$

$$
\begin{aligned}
&\le f(x)'(x - y_x^1) - \rho \left\| x - y_x^1 \right\|_{H(x)}^2 + f(y_x^1)'(y_x^1 - x), \text{ (using (2.2))} \\
&\le f(x)'(x - y_x^1) - \left\| x - y_x^1 \right\|_{H(z_x)}^2 + f(y_x^1)'(y_x^1 - x), \text{ (using Assum. 3)} \\
&= f(y_x^1)'(x - y_x^1) + f(y_x^1)'(y_x^1 - x), \text{ (using (2.3))} \\
&= 0.
\end{aligned}
$$

Thus we have proved the first statement of the lemma. We next show that the set of $WVIP$ solutions coincides with the set of minimizers of C_d^1. Let x^* be a $WVIP$ solution. Observe that $C_d^1(x) \geq 0$. Since $C_d(x^*) = 0 \leq C_d^1(x)$, for all $x \in K$, we have shown that $x^* \in \{z \in K \mid f(y_x)'(z - x) \leq 0\}$.

In what follows we show that $C_d(x) \geq C_d^1(x)$. The definitions of points y_x and y_x^1 imply that $C_d(x) = f(y_x)'(x - y_x)$ and $C_d^1(x) = f(x)'(x - y_x^1) - \rho \left\| x - y_x^1 \right\|_{H(x)}^2$ respectively. Therefore,

$$
\begin{aligned}
C_d(x) &\ge f(y_x^1)'(x - y_x^1) = f(x)'(x - y_x^1) - \left\| x - y_x^1 \right\|_{H(w)}^2, \text{ (using (2.3))} \\
&\ge f(x)'(x - y_x^1) - \rho \left\| x - y_x^1 \right\|_{H(x)}^2 = C_d^1(x).
\end{aligned}
$$

Since $C_d^1(x^*) \geq 0 = C_d(x^*)$, the previous inequality implies that $C_d^1(x^*) = 0$ and, hence, we have shown that any $WVIP$ solution minimizes the function $C_d^1(x)$. Next we show that every $\bar{x} \in \arg\min_{x \in K} C_d^1(x)$ is also a $WVIP$ solution. Suppose, the opposite is true, that is, for every such $\bar{x}$, there exists some y such that $f(y)'(y - \bar{x}) < 0$. The mean value theorem and the Jacobian similarity property imply that $f(\bar{x})'(\bar{x} - y) = f(y)'(\bar{x} - y) + \|y - \bar{x}\|_{H(\tilde{z})}^2 \geq f(y)'(\bar{x} - y) + \frac{1}{\rho}\|y - \bar{x}\|_{H(\bar{x})}^2$, for some $\tilde{z} \in [\bar{x}; y]$. Consider $\alpha = 1/\rho^2 \in (0, 1]$. Then for $y_\alpha = \bar{x} + \alpha(y - \bar{x}) \in K$, it follows that

$$
f(\bar{x})'(\bar{x} - y_\alpha) - \rho \left\| \bar{x} - y_\alpha \right\|_{H(\bar{x})}^2 = \alpha f(y)'(\bar{x} - y) - \frac{\alpha}{\rho}(\rho^2 \alpha - 1) \left\| \bar{x} - y \right\|_{H(\bar{x}))}^2 > 0.
$$

However, we argued above that $C_d^1(\bar{x}) = 0$, for all $\bar{x} \in \arg\min_{x \in K} C_d^1(x)$. This leads to a contradiction. ■

Remark. Under the Jacobian similarity property, the set of minimizers of gap function C_d^1 also coincides with the VIP solution set. This follows from the observation that for any VIP solution x^*, $f(x^*)'(x^* - y) - \rho \|x^* - y\|^2_{H(x^*)} \leq 0$. This, in turn, implies that $C_d^1(x^*) = 0$. It follows that every minimizer of C_d^1 is a VIP solution. In other words, under the Jacobian similarity property, the $WVIP$ solution set coincides with the VIP solution set.

Second approximation

In the previous subsection we considered a quadratic approximation of the dual gap function C_d. Although this approximation simplified the objective function in the dual gap function computation, it did not concern itself with the structure of the feasible region K. We next consider a polyhedral feasible region K of the form $\{x \mid Ax = b,\ x \geq 0\}$.

Assumption 4 *Matrix A has a full row rank.*

Assumption 5 *All $x \in K$ are primal non-degenerate, i.e., if $x_N = 0$ for $N \subseteq \{1, ..., n\}$, then A_B is of full rank where $B = \{1, ..., n\} \backslash N$.*

In the following development, we consider an approximation of the gap function C_d (similar to C_d^1) by also restricting the maximization problem over a Dikin ellipsoid rather than maximizing over the entire feasible region K. We denote a *Dikin ellipsoid* by $D(x) = \{y \in K \mid \|X^{-1}(y - x)\| \leq r\}$, where the matrix $X = \operatorname{diag}(x)$ and the constant $r \in (0, 1)$. We then define

$$C_d^2(x) = \max_y \left\{ f(x)'(x - y) - \rho \|x - y\|^2_{H(x)} \mid Ay = b, \left\| X^{-1}(y - x) \right\| \leq r \right\}.$$

Notice that under Assumptions 4 and 5, the function $C_d^2(x)$ is well defined for every x in the polyhedron K (Assumption 4 is sufficient for $C_d^2(x)$ to be well-defined in the interior of K). Moreover, the computation of point y_x^2, the maximizer in the definition of $C_d^2(x)$, when x lies in the interior of K, can be performed in polynomial time (see [57], [62]). The key properties that hold for $C_d^1(x)$, will also hold for $C_d^2(x)$.

Lemma 4 *Suppose that Assumptions 1-5 hold.*

1. *The cuts contain the $WVIP$ solution set:*
$\{z \in K \mid C_d(z) \leq C_d^2(x)\} \subseteq \{z \in K \mid f(y_x^2)'(z-x) \leq 0\}$.

2. *C_d^2 relates to the exact dual function as follows: $C_d^2(x) \geq 0$; $C_d(x) \geq C_d^2(x)$.*

3. *C_d^2 characterizes the solution set: $C_d^2(x^*) = 0$;*
x^* is a $WVIP$ solution if and only if $C_d^2(x^*) = 0$.

2.3 Dual function cuts in the generalized geometric framework

So far we have shown how to construct linear cuts that incorporate second order information about the VIP function and showed that these cuts always contain the VIP solution set. In this section, we introduce these cuts into the GGF ([39]) and provide complexity and convergence results for the GGF with these new types of cuts.

2.3.1 Preliminaries and key properties

Generalized geometric framework

General geometric framework (GGF) in [39] is a cutting plane method that incorporates several known cutting plane method studied in continuous optimization, such as, for example, ellipsoid method. A key notion in the GGF is that of a "nice" set. These "nice" sets are constructed at each iteration so that they have the following properties: (a) an approximation of a "nice" set and its center can be computed efficiently, (b) a "nice" sets contains the VIP (or $WVIP$) solution set, (c) the volume of the "nice" sets strictly decreases at each iteration. At each iteration, the GGF first computes a cut through the center of the "nice" set, and subsequently constructs a new "nice" set of smaller volume that contains the remainder of the feasible region as well as the VIP (or $WVIP$) solutions. We refer the reader to [39] for more details on the GGF and next summarize the basic features of the framework.

At iteration k, the GGF (see also Figure 2-1) maintains the following: 1) a "nice" sets P^k that are a compact convex subsets of $\mathbb{R}^n$ with nonempty interior and belong to the same class of sets; 2) a set $K^k = P^k \bigcap K$ that contains the solution set ($\dim(P^k) = \dim(K^k)$), and, finally, 3) the iterate $x^k \in \text{int } (P^k)$, which is the center of P^k. Each iteration introduces a cut Cut^k and a new set $P^{k+1} = Cut^k \cap P^k$. Initially set P^0 is chosen so that it contains the entire feasible region K and its volume is no greater than $2^{n(2L+1)}$. Furthermore, we assume that we can construct "nice" sets P^k so that their volumes strictly decrease, that is, $\text{Vol}(P^{k+1}) \leq b(n)\text{Vol}(P^k)$, for some constant $0 < b(n) < 1$. As a result, the volume $\text{Vol}(K^k)$ is non-increasing and converges to zero.

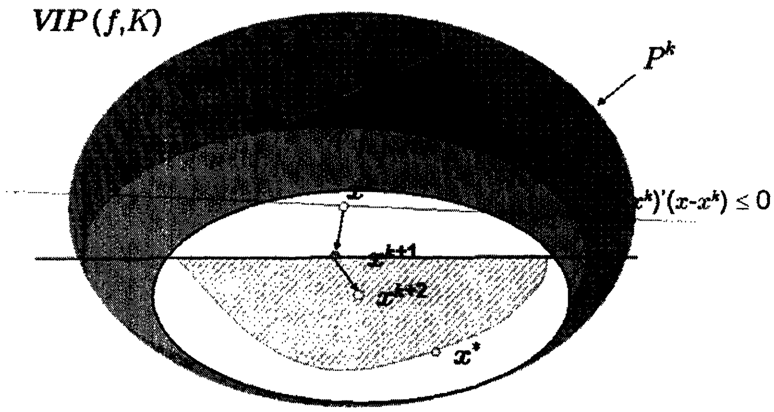

Figure 2-1: Generalized geometric framework

In [39] linear cuts with slopes $f(x^k)$ are considered. In this chapter, we introduce alternate choices for the slopes of linear cuts as discussed in Section 2. The extension of the GGF allows us to incorporate nonlinear cuts (see [51], Appendix B).

Approximate solutions

We begin the convergence analysis by describing the notion of an approximate solution in the context of a variational inequality as well as a weak variational inequality problem. The definitions we introduce use the gap function concepts described in Section 2.2.

Definition 5 For any $\varepsilon > 0$, a point $x^I \in K$ is an *ε-approximate VIP* solution if $C_p(x^I) \leq \varepsilon$, where C_p is the primal gap function.

Definition 6 For any $\varepsilon > 0$, a point $x^{II} \in K$ is an *ε-approximate $WVIP$* solution if $C_d\left(x^{II}\right) \leq \varepsilon$, where C_d is the dual gap function.

We introduce the following assumption:

Assumption 6 *f is a bounded function. That is, for some $M > 0$, $\|f(x)\| \leq M$, for any $x \in K$.*

Remark. Notice that we can also state scale-invariant versions of Definitions 5 and 6. For example, a point $x^I \in K$ is an *ε-approximate VIP* solution, if for any $\varepsilon > 0$, $C_p(x^I) \leq 2\varepsilon 2^{-l}M$. The constant M is defined in Assumption 6, and l is defined in Subsection 2.1.1. We can adjust the proofs in this section to also be applicable to the scale-invariant definition (see [39] for a more detailed discussion of these definitions).

At this point, it is natural to ask when Definitions 5 and 6 become equivalent. First, we notice that in Section 1 we introduced assumptions under which the variational inequality and weak variational inequality problems have the same solutions. However, these results do not directly translate into the equivalence of the respective approximate solutions. As a result, we next examine the relationship between approximate VIP and $WVIP$ solutions.

Assumption 7 *For some $\lambda > 0$, $d'\nabla f(x)d \leq \lambda d'd$, for all $x \in K$, $d \in \mathbb{R}^n$.*

The next two propositions establish a connection between approximate VIP and $WVIP$ solutions.

Proposition 1 *If the VIP problem function f is monotone, then an ε-approximate VIP solution x^I is also an ε-approximate $WVIP$ solution.*

Proof. This result follows from the definitions of approximate solutions and monotone functions. ■

Proposition 2 *Suppose that Assumptions 1 and 7 hold and L is the constant defined in Subsection 2.1.1. Then an ε-approximate WVIP solution is also a $\sqrt{2^{2L+2}\lambda\varepsilon}$-approximate VIP solution.*

Proof. Suppose x^{II} is an ε -approximate $WVIP$ solution, that is for all $z \in K$, $f(z)'(x^{II} - z) \leq \varepsilon$. Then an application of the mean value theorem implies that for some $y \in [x; z]$, $f(x^{II})'(z - x^{II}) = f(z)'(z - x^{II}) - \left\|z - x^{II}\right\|^2_{H(y)} \geq -\varepsilon - \lambda \left\|z - x^{II}\right\|^2$ (*). For $\alpha \in (0, 1]$, we now define point $z_\alpha = x^{II} + \alpha\left(x - x^{II}\right)$, for any $x \in K$. Then the convexity of set K implies that point z_α lies in the set K. Therefore, $f(x^{II})'(z_\alpha - x^{II}) = \alpha f(x^{II})'(x - x^{II})$, and so, $f(x^{II})'(x - x^{II}) = \dfrac{1}{\alpha} f(x^{II})'(z_\alpha - x^{II})$. Applying (*) for a choice of $z = z_\alpha$ shows that $f(x^{II})'(x - x^{II}) \geq \dfrac{1}{\alpha}\left(-\varepsilon - \lambda\alpha^2 \left\|x - x^{II}\right\|^2\right) \geq -\dfrac{1}{\alpha}\varepsilon - \lambda\alpha 2^{2L}$. This inequality is true for any choice of $x \in K$.

Setting $\bar{\varepsilon}(\alpha) = \dfrac{1}{\alpha}\varepsilon + \lambda\alpha 2^{2L}$, we conclude that x^{II} is an $\bar{\varepsilon}(\alpha)$−approximate VIP solution. In particular, the choice of $\alpha = \sqrt{\frac{\varepsilon}{\lambda 2^{2L}}}$ minimizes $\bar{\varepsilon}(\alpha)$. For this choice of α, $\bar{\varepsilon} = \sqrt{2^{2L+2}\lambda\varepsilon}$. Therefore, the point x^{II} is a $\sqrt{2^{2L+2}\lambda\varepsilon}$-approximate VIP solution. Notice that as $\dfrac{\varepsilon}{\alpha} \to 0$ and $\alpha \to 0$, $f(x^{II})'(x - x^{II}) \geq 0$, for all $x \in K$. ■

2.3.2 Convergence and complexity bounds

In this subsection we establish complexity bounds for the GGF with the cuts we introduced in Section 2. However, we first state a convergence result, that follows directly from the description of the GGF, the properties of dual gap function and Lemmas 1- 4.

Theorem 1 *Suppose that a WVIP satisfies Assumptions 1 and 2. Let $\{x^k\} \subset K$ be a sequence induced by the GGF with cuts $Cut(y^k, x^k)$, where y^k is defined for x^k as in (2.1). Every limit point of $\{x^k\}$ is a WVIP solution. Moreover, if instead we consider cuts $Cut(y^{1,k}, x^k)$ under an additional Assumption 3 or cuts $Cut(y^{2,k}, x^k)$ under additional Assumptions 3-5, with $y^{1,k}$ and $y^{2,k}$ being maximizers in the definitions of $C^1_d(x^k)$ and $C^2_d(x^k)$, respectfully, then the GGF converges to a VIP solution as well.*

To proceed in establishing complexity results, we need the following assumption.

Assumption 8 $2l \geq \log n + 1; L_1 = L + 3l + \log(\frac{M}{\varepsilon})$.

We introduce a contraction map $T : K \to K$ with $T(x) = x^* + 2^{-L_1+l+\log n}(x - x^*)$, where x^* is some solution of the VIP. This is a one-to-one map between the sets K and $T(K)$. Therefore, we can retrieve a point x from its image $y = T(x)$, as $x = x^* + 2^{L_1-l-\log n}(y - x^*)$.

Proposition 3 *Let* $\bar{k} = O\left(-\frac{nL_1}{\log b(n)}\right)$. *Assumptions 1, 6, 8 imply that there is a point* $y \in T(K) \cap P^{\bar{k}^c}$, *where* $P^{\bar{k}^c}$ *is the complement of* $P^{\bar{k}}$.

Proof. Observe that the volume of set P^k is at most 2^{-nL_1} in $\bar{k} = O\left(-\frac{nL_1}{\log b(n)}\right)$ iterations. Moreover, Assumption 1 implies that there is a point $z \in K$ such that the ball $S(z, 2^{-l})$ of radius 2^{-l}, centered at the point z , is contained in the feasible region K and $S = S(z, 2^{-l}) \neq K$. Therefore $T(S) \subseteq T(K)$ and

$$\text{Vol}(T(K)) > \text{Vol}(T(S)) = \text{Vol}(S(T(z), 2^{-L_1+\log n})) \geq \frac{(2^{-L_1+\log n})^n}{n^n} \geq \text{Vol}(P^{\bar{k}}).$$

Since Vol$(T(K))$ >Vol$(P^{\bar{k}})$, it follows that $T(K) \cap P^{\bar{k}^c} \neq \emptyset$. ■

Theorem 2 *Suppose Assumptions 1, 2, 6 - 8 are satisfied, then the GGF computes an* ε*-approximate* $WVIP$ *solution* $\bar{x}$ *in* $\bar{k} = O\left(-\frac{nL_1}{\log b(n)}\right)$ *iterations.*

Proof. As in the preceding proposition we first note that since the volume of set P^k is at most 2^{-nL_1} in $\bar{k} = O\left(-\frac{nL_1}{\log b(n)}\right)$ iterations, the GGF reaches a feasible solution in at most $\bar{k}$ steps. Furthermore, notice that Proposition 3 implies that there exists a point y that lies in $T(K) \cap P^{\bar{k}^c}$. Consequently,

$$\begin{aligned}
C_d(\bar{x}) &\leq C_d(y), \left(\text{since } y \in T(K) \cap P^{\bar{k}^c}\right) \\
&= f(y_y)'(y - y_y), \text{ (using the definition of } C_d(y)) \\
&= f(y_y)'(y - x^*) + f(y_y)'(x^* - y_y) \\
&\leq f(y_y)'(y - x^*), \text{ (since } x^* \text{ is a } WVIP \text{ solution)} \\
&= 2^{-L_1+l+\log n} f(y_y)'(z - x^*), \left(\text{setting } z = T^{-1}(y)\right) \\
&\leq 2^{-L_1+l+\log n} \cdot M \cdot 2^{L+1}, \text{ (using Assumption 6)}
\end{aligned}$$

$\leq \ \varepsilon$, (using Assumption 8).■

The previous theorem developed a complexity result for the weak variational inequality problem. We are now ready to prove a complexity result for the variational inequality problem.

Theorem 3 *Let $L_2 = L+3l+2\log(\frac{(M\lambda)^{0.5}2^{L+1}}{\varepsilon})$. Consider a VIP satisfying Assumptions 1 - 8. In $O\left(-\frac{nL_2}{\log b(n)}\right)$ iterations, the GGF computes an ε-approximate VIP solution.*

Proof. Theorem 2 implies that in $O\left(-\frac{nL_2}{\log b(n)}\right)$ iterations the GGF computes a point x, at which dual gap function $C_d(x) \leq \frac{\varepsilon^2}{\lambda 2^{2L+2}}$. Proposition 3 implies the result. ■

Since the exact dual gap function is often difficult to compute, the following results is useful for determining the stopping criterium.

Proposition 4 *In $\bar{k} = O\left(-\frac{nL_1}{\log b(n)}\right)$ iterations, $\bar{x}$ is reached for which $C_d^1(\bar{x}) \leq \varepsilon$ (or $C_d^2(\bar{x}) \leq \varepsilon$). Moreover, under Assumption 3, if $C_d^1(x) \leq \varepsilon$ (or $C_d^2(x) \leq \varepsilon$ and Assumptions 3 - 5 hold), then $C_d(x) \leq \rho^2\varepsilon$.*

Proof. Since for any $x \in K$, $C_d^i(x) \leq C_d(x)$, for $i = 1, 2$ (see Lemmas 3 and 4), it follows that in $\bar{k}$ iterations, $C_d^i(\bar{x}) \leq \varepsilon$ is satisfied.

Next we prove the second part of the proposition. Suppose, to the contrary, that for some x, $C_d^1(x) \leq \varepsilon$, while $C_d(x) > \rho^2\varepsilon$. Then for some $\bar{y}$, $f(\bar{y})'(x^k - \bar{y}) > \rho^2\varepsilon$. Consider a point $y_\alpha = (1-\alpha)x + \alpha\bar{y}$, for some $\alpha \in [0,1]$. We define the function $g(y_\alpha) \triangleq f(x)'(x - y_\alpha) - \rho \|x - y_\alpha\|_{H(x)}^2$. Then for some $z \in [x; \bar{y}]$,

$$\begin{aligned} g(y_\alpha) &= \alpha f(x)'(x-\bar{y}) - \alpha^2\rho \|x-\bar{y}\|_{H(x)}^2 = \alpha\left(f(\bar{y})'(x-\bar{y}) + \|\bar{y}-x\|_{H(z)}^2\right) - \alpha^2\rho\|x-\bar{y}\|_{H(x)}^2 \\ &> \alpha\left(\rho^2\varepsilon + \|\bar{y}-x\|_{H(z)}^2\right) - \alpha^2\rho\|x-\bar{y}\|_{H(x)}^2 = \alpha\rho^2\varepsilon + \alpha\|\bar{y}-x\|_{H(z)}^2 - \alpha^2\rho\|x-\bar{y}\|_{H(x)}^2 \\ &\geq \alpha\rho^2\varepsilon - \frac{\alpha}{\rho}\left(1-\alpha\rho^2\right)\|x-\bar{y}\|_{H(x)}^2. \end{aligned}$$

Letting $\alpha = \frac{1}{\rho^2}$, we observe that $g(y_\alpha) > \varepsilon$. Hence, $C_d^1(x) = \max_{y\in K} g(y) > \varepsilon$. This contradicts the assumption that $C_d^1(x) \leq \varepsilon$. Similarly we can show that $C_d^2(x) \leq \varepsilon$ implies that $C_d(x) \leq \rho^2\varepsilon$. Notice that if the property of Jacobian similarity holds for some constant ρ, then it also holds for all $\varrho \geq \rho$. Therefore, if we assume that $C_d(x) > \rho^2\varepsilon$, then $g(y_\alpha) > \varepsilon$, for every $\alpha \leq \frac{1}{\rho^2}$ (where y_α is defined above). Since for a sufficiently small α, point y_α lies inside the Dikin ellipsoid $D(x)$, it also holds that $C_d^2(x) > \varepsilon$, (i.e., $C_d(x) > \rho^2\varepsilon$ implies that $C_d^2(x) > \varepsilon$). ■

Example: (see [39]). All the methods that are special cases of the GGF can also be modified to incorporate the cuts we introduced in this chapter. Below we list some of these methods together with the respective descriptions of the volume reduction constants as well as the complexity bounds.

Method of centers of gravity: $b\,(n) = \frac{e-1}{e}$, $O\,(nL_2)$.

Ellipsoid Method: $b\,(n) = 2^{O\left(\frac{1}{n}\right)}$, $O\,(n^2L_2)$.

Method of inscribed ellipsoids: $b\,(n) = 0.843$, $O\,(nL_2)$.

Volumetric Center Method: $b\,(n) = const$, $O\,(nL_2)$.

2.4 Dual function based cuts: an application to the affine scaling method

In the previous section, we showed convergence and polynomial complexity for the GGF with cuts that are based on dual gap function. However, the task of finding a "nice" set P^k and its center x^k, at each iteration k, might be computationally difficult. Moreover, the complexity bounds for the methods in this framework tend to be tight in practice. For this reasons, in the remainder of this chapter, we will consider methods that are simpler to perform and, perhaps, as a result computationally more efficient. The version of affine scaling that we consider uses linear subproblem to find a direction towards next iterate (The convergence of this method in the context of convex optimization was first studied by Gonzaga and Carlos [28]). Other versions of affine scaling method for solving nonlinear optimization problems in the literature consider quadratic objectives in the direction finding subproblem (see [54], [61], [62]).

In the remainder of this section we assume that the feasible region is a polyhedron and that Assumptions 1 and 4 hold. We will also assume the following:

Assumption 9 *The Jacobian matrix of the problem function f is symmetric and positive semi-definite.*

2.4.1 Affine scaling method and convergence results

We first describe the method and then establish convergence. Because of Assumption 9, we can represent the problem as an optimization problem with the function f as the gradient of a convex objective function F (i.e., $f = \nabla F$).

The affine scaling algorithm (as in [28]) finds

$$d^k = \arg\min\left\{f\left(x^k\right)' d \mid Ad = 0,\ \|(X^k)^{-1}d\| \leq r\right\}$$

and for some $\theta \in (0,1)$ finds the step size $\alpha^k = \min\left(-\frac{f(x^k)'d^k}{\|d^k\|^2_{H(z^k)}}, \bar{\alpha}\right)$, where $z^k \in [x^k, x^k + \bar{\alpha}d^k]$, $\bar{\alpha} \in [\theta, \alpha_{max})$ and α_{max} is the maximum feasible step length. Notice that $d^k = d(x^k)$ as in the following formula:

$$d(x) = -r\frac{XPXf(x)}{\|PXf(x)\|}, \tag{2.4}$$

where P is projection matrix onto Null(AX).

Theorem 4 *([28]) Suppose that Assumption 1, 2, 7 - 9 are satisfied. Then any limit point of the sequence $\{x^k\}$ generated by the AS method is a VIP solution.*

Under the condition of strict complementarity, we can extend Theorem 4 to show that in fact the entire sequence converges to a solution when strict complementarity holds. First, we define strict complementarity.

Definition 7 A limit point $(\bar{x}, \bar{s})$ satisfies the property of *strict complementarity,* if $\bar{x}_i\bar{s}_i = 0$ and $\bar{x}_i + \bar{s}_i > 0$, $\forall i$.

Assumption 10 *Every limit point of the sequence $\{x^k, s(x^k)\}$ satisfies the strict complementarity property.*

We are now ready to establish the following result.

Theorem 5 *Under Assumptions 1, 2, 4, 5, 9, 10 the AS method converges to an optimal solution.*

Proof. For some limit point $\bar{s}$, we let $N = \{i \mid \bar{s}_i \neq 0\}$, $B = \{i \mid \bar{s}_i = 0\}$, $S^* = \{x \mid Xs(x) = 0,\ Ax = b,\ x \geq 0\}$, and $C_\delta = \{x \mid x_i \in [0, \delta)\ \forall i \in N\}$.

First, notice that Assumption 5 and the continuity of matrix AX^2A' imply that $s(x)$ is a continuous variable. Hence for a small enough $\delta > 0$, when $x \in C_\delta$ and $j \in N$, it holds that $|s_j| \geq \frac{1}{2}|\bar{s}_j| > 0$.

We next show that Assumption 10 implies that the set of limit points is discrete. Since the sequence $\{(x^k, s^k)\}$ is bounded, it has some limit point, say $(\bar{x}, \bar{s})$. Notice that every limit point $\bar{x}$ of the sequence $\{x^k\}$ induced by the AS method belongs to the set S^*. This set also contains all $x \in K$ satisfying the strict complementarity property. We will show that for some $\delta > 0$ there exists a neighborhood C_δ such that $C_\delta \bigcap S^* = \bar{x}$. Suppose, to the contrary, that $\forall \delta > 0$, $\exists x^\delta \in S^* \bigcap C_\delta$ such that $x^\delta \neq \bar{x}$. Then, since the point $\bar{x}$ is a vector of finite dimensions, for some $j \in N$, there exists a sequence $\{x^{\delta_n}\}$ with $x^{\delta_n} \in S^* \bigcap C_{\delta_n}$, $x^{\delta_n} \neq \bar{x}$, $x_j^{\delta_n} \to 0$, $x_j^{\delta_n} > 0$, and $|x_j^{\delta_n} - \bar{x}_j| < \frac{1}{n}$. Notice that the strict complementarity property implies that $s_j^{\delta_n} = 0$ whereas $\bar{s}_j \neq 0$. By continuity, however, $s_j^{\delta_n} \to \bar{s}_j$, which is a contradiction. Therefore, for any limit point of the sequence $\{x^k\}$, there exists a neighborhood whose intersection with the set of all limit points is a singleton.

At this point, we notice that conditions analogous to those in [62] are satisfied and the convergence of the sequence $\{x^k\}$ can be shown by contradiction. For the sake of brevity we omit this proof and refer the reader to [62] or [51] for more details. Once convergence of the sequence is established, Theorem 4 ensures that it converges to a solution. ■

2.4.2 Convergence of the AS method for asymmetric VIPs

In this subsection we examine the convergence of the AS method when applied to an affine asymmetric VIP with problem function $f(x) = Mx + c$, defined by a positive semi-definite matrix M. We note that the proofs for a symmetric matrix M ([28]) use the existence of potential a $F(x) = \frac{1}{2}x'Mx + c'x$ to show that $f(x^k)'d^k$ converges to 0. This subsequently implies that at the limit point, the KKT conditions are satisfied. Moreover, convexity implies that every limit point is a solution.

We observe that the direction d found in (2.4) is a decent direction for function $F(x) = \frac{1}{2}x'Mx + c'x$ $\left(\text{where } \nabla F(x) = \frac{M+M'}{2}x + c\right)$. Then

$$\nabla F(x)'d = -r\left(\frac{1}{2}\|PXf(x)\| + \frac{1}{2}\frac{(M'x+c)'XPX(Mx+c)}{\|PXf(x)\|}\right) \leq 0. \tag{2.5}$$

Hence d is a direction of descent when $(Mx+c)'XPX(Mx+c)+(M'x+c)'XPX(Mx+c) \geq 0$. In particular, suppose that the following assumption holds.

Assumption 11 *Matrix M is positive semi-definite and for all $x \in K$ the following condition holds:*

$$x'MXPXMx \geq 0. \tag{2.6}$$

We modify the step size to be $\alpha^k = \min\left(\frac{-d'(\frac{M+M'}{2}x+c)}{d'Md}, 1\right)$, which ensures that the sequence $\{F(x^k)\}$ is non-increasing. Since it is also bounded, it has a unique limit point, which we denote by $\bar{F}$. First, just like in the proofs in [28], the continuity of F implies that $\nabla F(x^k)'d^k \to 0$; therefore, under Assumption (11), $f(x^k)'d^k \to 0$. When this last condition holds, Lemmas 3 and 4 of [51] (or see similar results in [28]) imply that for any limit point $\bar{x}$, corresponding dual variables $\bar{s}$ are nonnegative and that the complementarity condition holds ($\bar{x}_i\bar{s}_i = 0$).

Finally, under Assumption 10, the sequence $\{x^k\}$ converges, and the rest of the results necessary to show convergence, follow as in the symmetric case. We summarize this discussion in the following theorem.

Theorem 6 *Suppose that for $f = Mx + c$ Assumption 11 holds. Then every limit point of the AS method is a VIP solution. Moreover, if Assumption 10 is satisfied, the method converges.*

Remark. For a nonlinear asymmetric VIP, we can use a potential based on the dual gap function to measure closeness to the solution set. For example, we can use $P(x) = f(x^*)'(x - x^*)$, where x^* is a VIP solution. Notice that $P(x) \geq 0$. Then we can show that the AS method converges, if for every VIP solution x^* and $\forall x \in \text{int}(K)$, the following condition holds: $f(x^*)'XPXf(x) \geq f(x)'XPXf(x)$.

A convergence rate under the Jacobian similarity property

In this subsection we establish a rate of convergence for the AS method. We assume that the problem function f satisfies the Jacobian similarity property (Assumption 3). An essential element underlying our approach here is the modification of the line search procedure in the AS method. In particular, we set the next iterate as $x^{k+1} = x^k + \alpha^k d^k$, where the direction d^k is chosen as in Subsection 2.4.1 and the "optimal" step size α^k is determined through a line search. That is, if for all $\alpha \in [0, 1]$, we define $x^k(\alpha) = x^k + \alpha\rho d^k$, where the constant ρ is the Jacobian similarity constant. Then we choose

$$\alpha^k \in [0,1] \text{ satisfying } (f(x^k(\alpha^k))'d^k) \cdot (\alpha - \alpha^k) \geq 0, \ \forall \alpha \in [0,1].$$

Notice that step size α^k equals the one in Subsection 2.4.1, scaled by $\frac{1}{\rho}$, where ρ is the Jacobian similarity constant.

In this subsection, we assume that the step sizes α^k are bounded away from zero. In particular,

Assumption 12 *For some $\alpha > 0$, $\alpha^k \geq \alpha$, for all k.*

Remark. Indeed some of the examples that we consider in Subsection 2.4.3, induce step sizes that satisfy this condition.

Proposition 5 *Consider the sequence of step sizes $\{\alpha^k\}$ as described above. For some $\bar{F}$ the following relation holds $F\left(x^{k+1}\right) - \bar{F} \leq \left(1 - \frac{\alpha^k}{2\sqrt{n}}\right)\left(F\left(x^k\right) - \bar{F}\right)$.*

Proof. We define an auxiliary point $y^k = x^k + \rho\alpha^k d^k$ and use a Taylor expansion to obtain

$$F(x^k) = F(y^k) + f(y^k)'(x^k - y^k) + \frac{1}{2}(x^k - y^k)'\nabla f(z^k)(x^k - y^k), \text{ (where } z^k \in [x^k; y^k])$$
$$\geq F(y^k) \text{ (using the definition of } y^k \text{ and the monotonicity of problem function } f.)$$

Moreover, the convexity of the objective function F and the previous inequality imply that

$$F(x^{k+1}) = F\left(\left(1 - \frac{1}{\rho}\right)x^k + \frac{1}{\rho}y^k\right) \leq \left(1 - \frac{1}{\rho}\right)F(x^k) + \frac{1}{\rho}F(y^k) \leq F(x^k).$$

Therefore, the sequence $\{F(x^k)\}$, with step sizes as defined above, is monotonically non-increasing. Since this sequence is also bounded, it has a limit point $\bar{F}$. Let $S = K \cap \{x \mid F(x) \leq \bar{F}\}$, where K is the feasible region. Notice that $S \neq \emptyset$, since every cluster point of the sequence $\{x^k\}$ belongs to this set. Using similar arguments as in [54], we can show that for sufficiently large k, $\min_{z \in S} \left\|(X^k)^{-1}\left(z - x^k\right)\right\| \leq \sqrt{n}$. In the following, we use this result to devise a feasible point in the direction finding subproblem. In particular, suppose that point $z^k \in S$ satisfies $\left\|(X^k)^{-1}\left(z^k - x^k\right)\right\| \leq \sqrt{n}$, then point $x = x^k + \frac{a}{\sqrt{n}}(z^k - x^k)$, with $0 \leq a \leq 1$, is feasible for the affine scaling direction finding subproblem. Therefore,

$$f(x^k)'d^k \leq \frac{a}{\sqrt{n}} f(x^k)'\left(z^k - x^k\right), \text{ for all } 0 \leq a \leq 1. \tag{2.7}$$

An application of the mean value theorem implies that

$$f(y^k)'\left(y^k - x^k\right) = \alpha^k \rho f(x^k)'d^k + (\alpha^k \rho)^2 d^{k'} \nabla f(\hat{z}) d^k, \text{ for some } \hat{z} \in \left[x^k; y^k\right].$$

Therefore, the definition of point y^k implies that

$$\alpha^k \rho d^{k'} \nabla f(\hat{z}) d^k \leq -f(x^k)'d^k. \tag{2.8}$$

Combining these results we obtain

$$\begin{aligned}
F(x^{k+1}) &= F\left(x^k\right) + \alpha^k f(x^k)'d^k + \frac{1}{2}(\alpha^k)^2 d^{k'}\nabla f(\tilde{z})d^k, (\text{ for some } \tilde{z} \in [x^k; x^{k+1}]) \\
&\leq F\left(x^k\right) + \alpha^k f(x^k)'d^k + \frac{\rho}{2}\alpha^{k2} d^{k'}\nabla f(\hat{z})d^k, \\
&\qquad \text{(using the property of Jacobian similarity)} \\
&\leq F\left(x^k\right) + \alpha^k f(x^k)'d^k - \frac{1}{2}\alpha^k f(x^k)'d^k, \text{ (using (2.8))} \\
&\leq F\left(x^k\right) + \frac{\alpha^k}{2}\frac{1}{\sqrt{n}} f(x^k)'\left(z^k - x^k\right), \text{ (using (2.7))} \\
&\leq F\left(x^k\right) + \frac{\alpha^k}{2}\frac{1}{\sqrt{n}}\left(F\left(z^k\right) - F\left(x^k\right)\right), \\
&\qquad \text{(using the convexity of the objective function } F) \\
&\leq \left(1 - \frac{\alpha^k}{2}\frac{1}{\sqrt{n}}\right) F\left(x^k\right) + \frac{\alpha^k}{2}\frac{1}{\sqrt{n}}\bar{F}.
\end{aligned}$$

Hence, $F\left(x^{k+1}\right) - \bar{F} \leq \left(1 - \frac{\alpha^k}{2}\frac{1}{\sqrt{n}}\right)\left(F\left(x^k\right) - \bar{F}\right)$. ■

This result allows us to obtain a rate of convergence for the AS method in this chapter. Next we illustrate that the sequence that this method induces is indeed convergent to an optimal solution.

Theorem 7 *Under Assumptions 3, 4, 9 and 12, the sequence $\{x^k\}$ generated by the AS method converges to an optimal solution.*

Proof. First, we observe that analogously to [54] (see also Appendix A), Assumptions 3 and 12 imply that there exists some constant $c > 0$, such that $F(x^k) - F(x^{k+1}) \geq cd^{k'}d^k$. Therefore,

$$\left\|x^k - x^s\right\| \leq \sum_{s+1}^{k}\left\|x^i - x^{i-1}\right\| = O\left(\sum_{s+1}^{k}\left|F\left(x^i\right) - F\left(x^{i-1}\right)\right|^{1/2}\right).$$

Assumption 12 and Proposition 5 imply that $F\left(x^{k+1}\right) - \bar{F} \leq \left(1 - \frac{\alpha}{2\sqrt{n}}\right)\left(F\left(x^k\right) - \bar{F}\right)$.

Hence,

$$F\left(x^{k+1}\right)-F\left(x^{k}\right)\leq F\left(x^{k+1}\right)-\bar{F}\leq\left(1-\frac{\alpha}{2\sqrt{n}}\right)\left(F\left(x^{k}\right)-\bar{F}\right),\text{ and as a result,}$$

$$\begin{aligned}\left\|x^{k}-x^{s}\right\| &\leq O\left(\sum_{i=1}^{k-s}\left(1-\frac{\alpha}{2\sqrt{n}}\right)^{i-1}\left|F\left(x^{s}\right)-\bar{F}\right|^{1/2}\right)\\ &=\frac{2\sqrt{n}}{\alpha}\left(1-\frac{\alpha}{2\sqrt{n}}-\left(1-\frac{\alpha}{2\sqrt{n}}\right)^{k-s}\right)O\left(\left|F\left(x^{s}\right)-\bar{F}\right|^{1/2}\right).\end{aligned}$$

Since the sequence $F\left(x^{s}\right)$ converges to $\bar{F}$, it follows that as $k, s\rightarrow\infty$, $\left\|x^{k}-x^{s}\right\|\rightarrow 0$. We conclude that $\left\{x^{k}\right\}$ is a Cauchy sequence and, therefore, it is a convergent sequence.

Moreover, since the function $s(x)$ is continuous in x, sequence $\{s^k\}$ is also convergent. Since $\alpha^k>\alpha$, $F(x^k)-\bar{F}\geq-\alpha f(x^k)'d^k+o(\|d^k\|^2)$. Then as shown in [51], $\|X^k s^k\|\rightarrow 0$. It follows that the limit of sequence $\{x^k\}$ satisfies the necessary conditions of optimality. The convexity of the objective function F implies the result. ■

Remark. Notice that in case of affine asymmetric VIPs, if α^k and F are defined as in Subsection 2.4.2, then all the results of this subsection still apply, moreover, to establish the rate of convergence in case of affine VIPs Assumption 12 is not necessary. Indeed, let $\bar{\alpha}=\min\left(\frac{-d'(\frac{M+M'}{2}x+c)}{d'Md},1\right)$ and let $\alpha^k=\bar{\alpha}/2$. Notice that $\nabla F(x^k+\alpha d^k)'d^k\leq 0\ \forall\alpha\in[0,\bar{\alpha}]$ and, therefore, it holds that $d^{k\prime}Md^k\leq-\frac{\left(\frac{M+M'}{2}x+c\right)'d^k}{\bar{\alpha}}$. Moreover, let $\bar{F}$ be minimizer of function F. Under Assumption 11, at every limit point of the AS method the function value would be $\bar{F}$ and $\bar{F}-F(x^k)\geq\alpha^k\nabla F(x^k)'d^k$. From these observations and assumptions, it follows that

$$\begin{aligned}F(x^{k+1}) &= F(x^k)+\alpha^k\left(\frac{M+M'}{2}x+c\right)'d^k+{\alpha^k}^2{d^k}'Md^k\\ &\leq F(x^k)+\alpha^k\left(\frac{M+M'}{2}x+c\right)'d^k-\frac{{\alpha^k}^2}{\bar{\alpha}}\left(\frac{M+M'}{2}x+c\right)'d^k\\ &= F(x^k)+\alpha^k(1-\frac{\alpha^k}{\bar{\alpha}})\left(\frac{M+M'}{2}x+c\right)'d^k\leq F(x^k)+\frac{\alpha^k}{2}\frac{(\bar{F}-F(x^k))}{\alpha^k},\end{aligned}$$

i.e., $F(x^{k+1})-\bar{F}\leq\frac{1}{2}(F(x^k)-\bar{F})$.

Moreover, for a general nonlinear VIP, the rate of convergence results would asymptotically hold when iterates are close to a solution. When close to a solution x^*, we can use potential $F(x)=f(x^*)'(x-x^*)+\frac{1}{2}(x-x^*)'H(x^*)(x-x^*)$ to establish a local rate of convergence as in Proposition 5. Let $\bar{\alpha}$ be a solution of one-dimensional VIP: $\alpha f(x+\alpha d)'d\leq 0$ on $[0,1]$. When x^k is close to the solution the following hold:

- $x^k + \frac{\alpha}{\sqrt{n}}(x^* - x^k) \in D(x^k), \ \forall \alpha \in [0,1]$;
- $\nabla F(x^k)'d = f(x^k)'d + o(\|x - x^*\|^2)$;
- $d^{k'} H^* d^k \leq -\frac{\nabla f(x^k)' d^k}{\alpha \rho}, \ \forall \alpha \in [0, \bar{\alpha}]$;
- $\bar{F} - F(x^k) \geq f(x^k)'(x^* - x) \geq \frac{\sqrt{n}}{\alpha} f(x^k)' d^k, \ \forall \alpha \in [0,1]$.

Based on these observations, we estimate

$$\begin{aligned} F(x^{k+1}) &= F(x^k) + \alpha^k \nabla F(x^k)' d^k + \frac{{\alpha^k}^2}{2} d^{k'} H^* d^k \\ &\leq F(x^k) + \alpha^k \nabla F(x^k)' d^k - \frac{\alpha^k}{2\rho} \nabla f(x^k)' d^k \\ &\leq F(x^k) + \alpha^k \frac{2\rho - 1}{2\rho} f(x^k)' d^k + o(\|x - x^*\|^2) \\ &\leq F(x^k) + \alpha^k \frac{2\rho - 1}{2\rho} \frac{\tilde{\alpha}}{\sqrt{n}} f(x^k)'(x^* - x^k) + o(\|x - x^*\|^2) \\ &\leq F(x^k) + \frac{\alpha}{\sqrt{n}} (\bar{F} - F(x^k)) + o(\|x - x^*\|^2), \ (\text{letting } \tilde{\alpha} = \frac{2\rho - 1}{2\rho}) \end{aligned}$$

i.e., $F(x^{k+1}) - \bar{F} \leq (1 - \frac{\alpha}{\sqrt{n}})(F(x^k) - \bar{F}) + o(\|x - x^*\|^2)$.

Finally observe that convergence arguments of Theorem 7 extend to asymmetric VIPs.

2.4.3 The AS method with cuts and computational results

Our goal in this section is to examine computationally the performance of cuts we introduced in this chapter. For this reason, we consider the affine scaling method as well as several variations that incorporate the cut ideas from Section 2. To test the methods in this chapter, we chose several randomly generated instances of affine variational inequality problems. In most of the examples the matrix M is symmetric. However, we also tested asymmetric examples chosen so that both M and M^2 are positive-definite. This last condition is similar to the condition necessary for convergence of generalized descent framework for asymmetric VIPs as described in [40]. In particular, we considered variational inequality problems of the following format,

$$\text{Find } \ x^* \in K \text{ such that } \ f(x^*)'(x - x^*) \geq 0, \quad \forall x \in K,$$

with (1) a problem function $f(x) = Mx - c$ defined by a positive semi-definite $n \times n$ matrix M and either (i) M is symmetric or (ii) M^2 is positive definite, and a vector

$c \in \mathbb{R}^n$, and (2) a polyhedron $K = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ defined by an $m \times n$ matrix A and a vector $b \in \mathbb{R}^m$. We generated the elements of matrix $M \in \mathbb{R} \times \mathbb{R}$ randomly so that M is diagonally dominated and M and M^2 are positive definite. The feasible region was generated in the form $\{x \mid x \geq 0,\ a'x \leq 1\}$ where $a \in \mathbb{R}^{n+}$ was some random positive vector. The initial strictly feasible point was generated as $x^0 = .5a^{-1}$ where $a^{-1} = (a_1^{-1}, \cdots, a_n^{-1})$. Asymmetric examples where generated similarly. In addition, we used several small-dimensional examples in which the cycling behavior of Frank-Wolfe algorithm could be easily observed. We implemented all the methods using MATLAB version 6.1, on a personal computer with a Dual Xeon 1.5GHz processor and 1GB RAM of memory. Finally, we computed the CPU time using MATLAB's build-in function. As a stopping criterion, we used as tolerance level $\varepsilon = 10^{-4}$.

Symmetric VIPs

We first study the performance of AS method for symmetric VIPs. In Table 2.1, we compare the performance of the affine scaling method introduced in this chapter with the performance of the Frank-Wolfe method ([22]). We observe that in all the examples we studied, the affine scaling method in this chapter fixes the zigzagging behavior of the Frank-Wolfe method. Moreover, in most of the examples, the affine scaling method computes a solution in less CPU time. However, in one of the examples, the Frank-Wolfe method computed the solution faster (in terms of CPU time) than the affine scaling method. In this example, the Frank-Wolfe method did not zigzag. The solution of the variational inequality problem lied at a corner point of the polyhedral feasible region, and the Frank-Wolfe method computed the solution in one step (that is, by solving a single linear optimization subproblem). It is worth noting that a possible reason for the faster convergence of the Frank-Wolfe method in this example, is that as it solved a single linear optimization subproblem using MATLAB's built-in optimization solver, it relied on the speed of this implementation. On the other hand, in the same example, the affine scaling method applied a sequence of much simpler steps. We believe that a better implementation of the affine scaling method will yield a comparable performance even in this example. In conclusion, we observed that in most of the examples we generated, the affine scaling method outperformed the Frank-Wolfe method.

m	n	FW iterations	FW time	LAS iterations	LAS time
8	7	2	0.0628	10	0.028
10	6	>3000	106.6596	11	0.0315
31	30	18	0.8967	137	1.1397
51	30	>3000	>171	46	0.7505
43	40	2	0.0691	21	0.3962
101	100	>3000	171.244	221	22.0946
171	100	>300	>300	803	252.9649
111	110	269	139.112	801	99.2196

FW Frank-Wolfe Algorithm
LAS long step Affine Scaling Algorithm

Table 2.1: LAS method vs. FW method

Moreover, in Table 2.2, we compare the performance of the long step affine scaling method we introduced in this chapter (LAS) with that of two other affine scaling methods: i) QLAS, a quadratic approximation long step affine scaling method (see [62]), ii) DLAS, a long step affine scaling method that considers a quadratic approximation of the objective further using a diagonal approximation of the Jacobian matrix. We chose the long step versions of these methods since we noticed that within the family of affine scaling methods the long step versions perform the best computationally. The two quadratic approximation methods (QLAS and DLAS) perform similarly computationally. Furthermore, in most of the examples we generated, the affine scaling method introduced in this chapter outperformed both of these methods in CPU time. We attribute this partly to the simplicity of each iteration. Nevertheless, in two examples the affine scaling method (LAS) performed worse. Even in these two examples, we drastically improved the performance of the method (LAS) when we incorporated cuts (see Table 2.3). This new method significantly outperformed the two quadratic approximation methods. In what follows, we will discuss this in further detail.

To further improve the computational results in this chapter, we incorporate the cut ideas we discussed in Section 2 into the affine scaling method. The motivation

m	n	LAS iterations	LAS time	QLAS iterations	QLAS time	DLAS iterations	DLAS time
8	7	10	0.028	10	0.0685	9	0.0667
11	10	9	0.0468	10	0.0859	10	0.0904
31	30	137	1.1397	12	1.3045	11	1.2324
51	30	46	0.7505	15	1.4003	12	1.2455
43	40	21	0.3962	16	3.2849	16	3.4243
166	60	153	41.5669	100	92.3103	100	95.1253
154	77	100	22.3608	95	189.2372	100	201.0876
101	100	221	22.0946	20	87.9	24	101.2705
171	100	803	252.9649	17	79.537	49	234.6041
111	110	801	99.2196	20	97.3761	15	69.733

LAS long step Affine Scaling Algorithm
QLAS LAS with a quadratic approximation of the objective in the subproblem
DLAS LAS with a diagonalized approximation of the objective in the subproblem

Table 2.2: LAS method vs. QLAS and DLAS methods

in this comes from the observation that methods utilizing cuts often provide better complexity results in theory as well as in practice.

The AS Method with Cuts

1. Start with a strictly feasible point x_0, feasible region $K^0 = K$, tolerance $\varepsilon > 0$, and constant $r \in (0,1)$.
2. At iteration k:
   (a) Find $y^k \in K$ such that $F(y^k) \leq F(x^k)$.
   (b) $d^k = \arg\min\{f(y^k)'d \mid Ad = 0,\ \|(Y^k)^{-1}d\| \leq r\}$.
   (c) Choose step size α^k s. t. $(\alpha - \alpha^k)f(y^k + \alpha^k d^k)'d^k \geq 0$, for all $\alpha \in (0, \alpha^k_{\max})$, where $\alpha^k_{\max} \geq 1$.
   (d) $x^{k+1} \leftarrow y^k + \alpha^k d^k$.
   (e) Update $K^{k+1} = K^k \cap Cut(y^k, x^k)$.
3. Stop when $|f(x^k)'d^k| \leq \varepsilon$.

In what follows, we compare the long step affine scaling method (LAS) with several special cases of the AS method with cuts we just described. These special cases include the following:

(1) **Simple cuts.**

Set point $y^k = x^k$ and introduce cut $Cut(x^k, x^k) = \left\{x \mid f\left(x^k\right)'\left(x - x^k\right) \leq 0\right\}$ (LASC).

(2) **Cuts based on the dual gap function.**

Suppose that Assumption 3 holds. We introduce cut $Cut(y^k, x^k) = \left\{x \mid f\left(y^k\right)'\left(x - x^k\right) \leq 0\right\}$ with

(a) $y^k = \arg\max_{y \in K, y \geq x^k} \left\{ f(x^k)'(x^k - y) - \rho \left\| x^k - y \right\|^2_{H(x^k)} \mid Ay = b, \| (X^k)^{-1}(y - x^k) \| \leq r \right\}$ (LASGs).

(b) $y^k = \arg\max_{y \in K} \left\{ f(x^k)'(x^k - y) - \rho \left\| x^k - y \right\|^2_{H(x^k)} \mid Ay = b, \| (X^k)^{-1}(y - x^k) \| \leq r \right\}$ (LASGD).

m	n	LAS		LASC		LASGs		LASGD	
		iterations	time	iterations	time	iterations	time	iterations	time
8	7	10	0.028	10	0.0325	6	0.0586	8	0.052
11	10	9	0.0468	9	0.0302	6	0.066	7	0.0678
31	30	137	1.1397	24	0.2985	8	0.9622	17	1.9883
51	30	46	0.7505	24	0.5143	9	0.7311	13	1.2917
43	40	21	0.3962	21	0.6716	10	0.8234	18	2.2375
166	60	153	41.5669	161	33.2601	12	4.1987	181	61.9024
154	77	100	22.3608	104	25.89	13	4.6085	133	104.8659
101	100	221	22.0946	25	2.9223	12	13.4612	89	392.7285
171	100	803	252.9649	40	15.8761	10	13.4899	50	241.7931
111	110	801	99.2196	33	4.659	13	12.9911	26	108.0049

LAS long step Affine Scaling Algorithm
LASC LAS with cuts Cut(x,x)
LASGs LAS with cuts Cut(y,x), where y is found within { z | Az >=a Ax , with 0<a<=1}
LASGD LAS with cuts Cut(y,x), where y is found within Dikin ellipsoid

Table 2.3: AS method vs. AS methods with cuts

Table 2.3 summarizes the computations that compare the various versions of the affine scaling method of this chapter. We notice that the two best versions in terms of CPU time are the method that uses simple cuts (LASC) as well as the method that uses cuts determined via a gap function where the direction of the cut is found within a restricted feasible region (LASGs). Moreover, Table 2.3 demonstrates that these two versions compute a solution in a comparable or even less number of iterations than the quadratic approximation affine scaling methods. Nevertheless, in terms of CPU time, both methods with cuts are faster. Among the versions of the method with cuts determined via a gap function, the method where the direction of the cut is found within a restricted feasible region (LASGs) has consistently the least number of iterations. In conclusion, both LASC and LASGs outperform considerably the affine scaling method without cuts, both in terms of number of iterations and in terms of CPU time.

Asymmetric VIPs

We also generated several examples of asymmetric affine VIPs. Table 2.4 compares the long step AS method (LAS) with the same methods as we considered for symmetric problems in the previous subsection. When constructing examples, we generated matrix M so that M and M^2 are positive definite, which is a rough approximation of the conditions in Assumption 11. We also checked whether condition (2.6) is satisfied at each iterate, i_{A11} in the table indicates the number of iterates at which this condition was satisfied for the LAS.

		LAS			LASC		LASGs		FW		QLAS	
m	n	i	t	i_{A11}	i	t	i	t	i	t	i	t
101	100	>3000	276	84	20	2.1	14	17	3000	263	23	133
171	100	1024	280	1024	29	7.6	11	13	>212	>300	17	79.1
123	41	89	8.5	89	112	8.4	12	1.6	2	0.79	67	12.4
63	41	942	27	942	61	1.7	31	2.7	>926	>300	29	9.5

i iterations
t CPU time
i_{A11} number of iterations at which Assumption 11 holds

Table 2.4: Asymmetric VIPs

We observe that in the most examples one of the methods with linear cuts (LAS or LASGs) outperformed the other methods in both CPU time and the number of iterations. We also observe that when condition (2.6) is satisfied at all the iterates, the AS method without cuts is converging, while when this condition is not satisfied, the method seems to cycle. This emphasizes the sufficiency of Assumption 11 for the convergence of the AS in asymmetric problems. Finally, methods with cuts are likely to substantially improve the convergence properties of the AS method, as was also displayed in the case of symmetric problems.

Summary of computational results

Below we summarize our learnings from the computational experiments we performed.

1. All the versions of the affine scaling method we introduced in this chapter fix the zigzagging behavior of the Frank-Wolfe method.

2. The AS method converges in case of both symmetric and asymmetric affine monotone VIP; for asymmetric VIPs it converges when Assumption 11 holds.

3. The cuts often speed up the AS method.

4. In two examples the Frank-Wolfe method performed better than the affine scaling method. These were an examples where the Frank-Wolfe method did not zigzag, but rather the Frank-Wolfe method found the solution in one step through the solution of a linear optimization subproblem. We attribute this to the quality of the MATLAB's built-in linear optimization solver. We believe that a better implementation of the affine scaling method will also yield comparable results in terms of CPU time, even in this example.

5. In theory, the affine scaling method will perform in the worst case similarly to the Frank-Wolfe method. Nevertheless, in most cases in practice, we believe that the affine scaling will perform better.

6. After comparing all the versions of the affine scaling method we considered in this chapter, we conclude that the LASC and LASGs methods perform better in terms of CPU time. Moreover, in terms of number of iterations, the LASC method performs a similar number of iterations while the LASGs method performs fewer iterations than the quadratic approximation affine scaling methods.

7. The LASC and LASGs versions of the affine scaling method outperform the affine scaling method (LAS) without cuts both in terms of number of iterations and in terms of CPU time.

8. In particular, the LASGs method performs fewer iterations than all the other versions of the affine scaling method we considered in this chapter. Furthermore, it has the best or second best CPU time.

9. As the dimension of the problem grows, the LASC and LASGs versions of the affine scaling method are likely to outperform the Frank-Wolfe method, the

quadratic approximation affine scaling methods we considered, and finally, the affine scaling method without cuts.

We would like to note that the theoretical convergence properties of the general affine scaling method with cuts are similar to those of the affine scaling method without cuts. For the sake of brevity we do not include this discussion in the chapter.

2.5 Conclusions

In this chapter, we have introduced "smarter" linear cuts whose directions are based on the dual gap function associated with variational inequality problems. We established complexity results for the GGF with these cuts. To investigate the computational effect of these cuts, we applied them to a version of the AS method for VIPs. This version of the AS, at each step a direction was computed by solving a subproblem with a linear objective within a Dikin ellipsoid. Under some additional assumptions imposed upon the step sizes (and, in case of asymmetric VIPs, on the region), we established the rate of convergence of this method for monotone variational inequality problems with no cuts present. In our computational experiments (for both symmetric and asymmetric VIP examples, with the latter satisfying additional assumption), we observed that even the affine scaling method without cuts outperformed (in terms of CPU time) the Frank-Wolfe algorithm as well as variations of the affine scaling method that use a quadratic approximation of the objective in the direction finding subproblem. Furthermore, the affine scaling method with cuts considerably reduced the number of iterations as well as the CPU time for larger dimensional examples as compared to other methods. For the asymmetric VIP examples, we observed that the method converges when the condition (2.6) holds at each iterate. Although more extensive computational testing is needed, preliminary testing seems to indicate that these methods might perform well in practice.

Chapter 3

Multi-period Models with Capacities in Competitive Supply Chain

3.1 Introduction and position within literature

This chapter studies the oligopolistic competition of capacitated sellers in a multi-period supply chain environment. We consider n sellers (suppliers) competing for the orders from a single buyer (retailer). At each period the buyer can place an order at the prices announced by the sellers. We assume that all the agents are profit maximizers, the buyer can carry inventory and faces dynamic demand.

Instances of dynamic oligopolistic competition among capacitated sellers can be found in many industries. For example, electrical utilities are serviced by finite numbers of power generators with limited capacities, from whom the utilities buy on a periodic basis. Similarly, in other natural resource industries, several sellers with limited capacities form oligopolies and compete for the orders from buyers. Generally, commodity or near-commodity exchanges with a finite number of sellers and large-volume buyers can supply examples of the kind of oligopolies that can fit our model. With the proliferation of internet-based marketplaces, this type of exchanges became an increasingly powerful presence in B2B transactions. Many B2B marketplaces are buyer-oriented and, in addition to lowering costs, leadtimes and allowing greater transparency in the supply chain, add value to buyers through (i) aggregating demand and thus creating artificially large-volume buyers (FOB.com) and (ii) introducing a broader supplier base. The competition among suppliers is usually represented through catalog auctions (catalogs list several suppliers allowing a buyer to comparison shop, e.g., ChemConnect, Sciquest) or through reverse auctions (suppliers offer bids for a particular order, e.g., Freemarkets).

Many transactions that are facilitated by e-marketplaces are of a one-time nature, whereas companies still negotiate for long term contracts. Nevertheless, multi-period demand can be satisfied through contracts made dynamically in a competitive environment, where both demand and suppliers' characteristics can change over time. We believe that one of the main contributions of this work lies in considering the dynamic aspect of oligopolistic supply chain competition.

Recently many researchers in the Operations Management (OM) field have been studying **supply chain competition**. Some examined equilibrium and the algorithms to compute it in a many-supplier/many-retailer models (see, for example, Nagurney, Dong and Zhang [45]) or in a one-supplier/many-retailer case (see, for example, Bernstein and Federgruen [3]). However, the majority of the models focus on competition between **a single buyer and a single seller.** This is sometimes easily extendable to the cases of multiple buyers or sellers. Cachon [9] studies a periodic review model of an inventory game between a seller and a buyer, where the demand is an exogenous stochastic process, both the buyer and the seller can carry inventory and experience lead times. [9] finds the equilibrium of the underlying game and studies how to coordinate this chain. The paper also surveys models of competitive supply chain. Similar models are also presented in Cachon and Fisher [8] and Caldentey and Wein [12]. Various coordination contracts in a one-buyer/one-seller case can also be found in Cachon and Lariviere [7] and in Lariviere [35]. Coordination using option contracts was first introduced in Barnes-Shuster, Bassok and Anupindi [1] in the case when a buyer has an option of purchasing the product at a market price. [1] showed that an equilibrium options contract will coordinate the chain; moreover, many of the commonly studied coordination contracts are special cases of options contracts. **Option contracts** in a model with a single capacitated seller and a single buyer were also studied by Wu, Kleindorfer, Sun and Zhang [60] and Spinler [53]. The first paper studied equilibrium prices, while the second paper extended the model to include state dependent options. Even though most models in the literature present either periodic review or static setting, **dynamic competition** has also been studied in continuous time setting (see Dockner et al. [20] and Jørgensen and Kort [33]).

So far we reviewed general trends in modelling one supplier/one buyer supply chain competition. **Oligopolistic competition** has been studied by economists starting with Cournot and Bertrand and there exists a substantial body of literature related to this subject. We refer the interested reader to Fudenberg and Tirole [23], [24] and Tirole [55] for more information and further references. In the context of OM, the competition of **several suppliers for the orders from a single buyer** was studied in Jin and Wu [31], which examined catalog auctions under perfect and asymmetric information. Cachon and Harker [10] studied competition between two firms (sellers) in scale economies. Specific dynamic pricing mechanisms among multiple sellers and a single buyer have also been studied in the literature: Gallien and Wein [25] considered a Freemarkets-like reverse auction; Chen, Jamakiraman, Roundy, and Zhang [14] considered generalized Vickrey auctions and accounted for transshipment costs; Beil and Wein [2] studied parameter estimation in a generalized multi-attribute auction; and Ertogral and Wu [21] propose a bargaining game that coordinates multi-seller supply chain. We also would like to mention here a survey by Cachon and Netessine [11] that provides a useful background on game theory in the context of supply chain competition.

The problem of a capacitated oligopoly has been examined extensively in the economics literature, beginning with Edgeworth [15] who studied models of price competition (see also Levitan and Shubik [36]). It was noted already by Edgeworth that when capacities are incorporated in an oligopolistic price competition model, finding equilibrium policies seizes to be as straightforward as in the case of pure price competition (Bertrand competition). [36] showed existence of a mixed strategy equilibrium. There are several versions of such existence results in the literature depending on the allocation rule. It was furthermore shown, that in several situations the existence of pure strategy equilibria can be justified: (1) if the game is modelled as a multi-period game with a stream of customers with a single unit demand, (2) if other parameters are present in the model (for example, a market price), or (3) if a certain relationship between the demand and the capacities holds. We note that the mixed strategy equilibrium for a static Edgeworth competition is derived in Kreps and Scheinkman [34]. In this thesis, we are interested in pure strategy equilibria, as these strategies are more intuitive and can be explained and implemented in practice. The collusion and trigger-price strategies in the repeated games context have also been studied by economists for oligopolies with capacitated suppliers (see [23], [24] for further references) and are beyond the scope of this work.

In the OM literature, the competition of capacitated sellers was studied by Wu, Kleindorfer, and Zhang [60] in a static environment. The authors consider a problem with a single buyer, n sellers. The buyer enters into option contracts with the sellers and each contract fixes a price and a capacity. Then the buyer observes the market price and demand and decides to exercise some of the contracts or to make purchases at a market price. The paper shows that option contracts coordinate this supply chain. A similar model is also studied in Martínez-de-Albéniz and Simchi-Levi [42], who focus on the properties of the equilibria of this model. Our study of oligopoly differs from the above mentioned work [42] and [60] in the following way.

- Our main model does not incorporate a "market" supplier. Therefore, we need to study contracts other than options to coordinate the chain.

- We consider dynamic competition in discrete and continuous time.

- We consider a profit-maximizing retailer (unlike [42] that considers a cost-minimizing retailer or [60] that considers a utility-maximizing retailer).

- We consider expandable capacities.

- Unlike [42], in this work, the capacity of suppliers is fixed and bounded from above.

The properties of the retailer's profit function in a discrete multi-period setting are also studied in Martínez-de-Albéniz and Simchi-Levi [41], where the retailer is a cost-minimizing buyer of capacity options from several suppliers. Even though some of the properties of the retailer's problem in this thesis are similar to those in [41], our focus is on the study of the equilibrium properties as opposed to an in-depth study of the retailer's problem. Moreover, this work also considers a continuous time model of competition.

Our model is also related to the **inventory theory literature**, as we consider a two-echelon supply chain, with the retailer constituting the second echelon and n suppliers forming the first echelon. The strategy of each of the players is determined through a general dynamic program, that is similar, for example, to the one considered in Chan, Simchi-Levi, and Swann [13] (when time is discretized) and in Bertsimas and Paschalidis [5] (when time is continuous). There are two major differences with these papers. First, as a result of competition among suppliers, the prices they announce will depend on the total amount ordered by the retailer, i.e., these prices do not remain constant in time and are piecewise linear in the total amount ordered by the retailer. Moreover, further differences in the equilibrium structure are imposed by the restrictions on the suppliers' capacities.

We believe that our contributions can be summarized as follows:

1. We study **dynamic** competition with **multiple capacitated** suppliers competing for an order from a single retailer (a) in continuous and (b) in discrete time setting. We characterize equilibrium policies for these settings.

2. We study **contracts** to coordinate this dynamic supply chain.

3. We study **option contracts** in a multi-period, multi-supplier setting.

4. We consider some extensions of the model including more general supplier cost structures, expandable capacities, and stochastic demand at the retailer.

The remainder of this chapter is organized as follows. Section 3.2 describes the decentralized and centralized models as well as the properties of equilibrium policies in the continuous time case. Section 3.3 describes the model and equilibrium policies in discrete time case. Section 3.4 considers coordination mechanism and compares decentralized and centralized solutions. Section 3.5 describes a modified model that incorporates options. Section 3.6 provides numerical examples, and Section 3.7 summarizes some extensions. We conclude in Section 3.8 outlining contributions and further research possibilities.

3.2 Continuous time model

3.2.1 Introduction

In this section we introduce and analyze general models of dynamic competition in a continuous time setting. In particular, we focus on a supply chain setting. We model the competition among n suppliers for an order from a single retailer over the time interval $[0, T]$ (see also Figure 3-1). We assume that competition is for a single homogeneous product[1]. The retailer can carry inventory I and faces continuous consumers' demand rate D. There are consumers' prices $p(D)$ that correspond to this demand rate and, therefore, are exogeneous to the system. Each supplier i produces a unit of product at a cost s_i and has a cap on the rate of input that equals K_i. We assume that suppliers' costs s_i and capacities K_i are constant over time. Supplier i charges the retailer a price w_i per unit of product. As a result the retailer orders at the rate q_i from this supplier. Without loss of generality, we can assume that the suppliers are indexed in increasing order of their costs. In the model we omit lead times as well as a possible production capacity at the retailer. Nevertheless, the results can be extended to include these features. Table 3.1 below summarizes the description of parameters and variables of the system at time t.

[1]As we mentioned earlier the product could be a commodity or near-commodity, i.e. a homogeneous or a highly substitutable one. Electricity, raw materials, chemicals, as well as milk, flour, coffee, paper, staples, etc., can be be considered as such products.

We note here that similar notation will be used later in the discrete time model. We will use superscript t to denote time in the discrete time case. Moreover, wherever it is unambiguous, we will omit the time parameter.

$(s_1, \ldots, s_n) = \mathbf{s}(t) = \mathbf{s} \in \mathbb{R}^{n+}$	vector of suppliers' costs, constant over time
$(w_1, \ldots, w_n) = \mathbf{w}(t) \in \mathbb{R}^{n+}$	vector of suppliers' prices
$(K_1, \ldots, K_n) = \mathbf{K}(t) = \mathbf{K}$	vector of upper bounds on replenishment rates
$K^k = \sum_{i=1}^{k} K_k$	total capacity of first k suppliers
$(q_1, \ldots, q_n) = \mathbf{q}(t) \in \mathbb{R}^{n+}$	vector of rates at which the product is ordered from the individual suppliers
$q(t) = \sum_{i=1}^{n} q_i(t)$	rate at which the product is ordered by the retailer
$I(t)$	retailer's inventory at the beginning of period t
$S(t)$	retailer's base stock level for period $t+1$
$D(t)$	consumers' demand, a random variable with cdf F, pdf f
$ED(t) = E(D(t))$	expected consumers' demand
$h(t) = h$	retailer's holding cost, constant over time
$b(t) = b$	retailer's costs of a lost sale, constant over time
$p(D)$	retailer's price, a function of consumers' demand D
$R(y, x) = p(x)\min(y, x)$	retailer' s revenue, a function of consumers' demand x and supply y at the retailer
$r(y) = R(y, ED)$	retailer's revenue for a given expected demand
$\pi_r(\mathbf{q}, S, t \mid \mathbf{w}, I)$	retailer's profit given initial inventory I and suppliers' prices $\mathbf{w}$ at time t
$\pi_i(w_i \mid \mathbf{w}_{-i})$	supplier i's profit
$DP^t(I \mid \mathbf{w})$	retailer's profit-to-go starting at time t, inventory I
$\pi^c(\mathbf{q} \mid I)$	centralized chain's profit given initial inventory I
$I^+(t)$	$= \max(I(t), 0)$
$I^-(t)$	$= \max(-I(t), 0)$
$\mathbf{w}_{-i}(t)$	$= (w_1, ..., w_{i-1}, w_{i+1}, ...w_n)$

Table 3.1: Notation

The continuous time setting is interesting to study since it allows us to gain an insight into the structure of equilibrium policies and can approximate some applications where there is high throughput and demand and competition among suppliers is dynamic. For example, the continuous time setting can be used to model competition of electric utilities/generators. In Section 3.3 we will also consider a discrete time setting. A discrete time model allows us to model more realistic settings (i.e., impose more realistic assumptions, allow parameters to change more slowly when the time intervals in the discretization are larger) as well as to get more efficient schemes for computation.

Decentralized vs. centralized models

The primary model we study will assume that each of the agents tries to maximize his/her individual profit. That is, each agent's strategy is locally controlled. Such systems are referred to as *decentralized* or *under decentralized control*, and at equilibrium, each agent's strategy is locally optimal.

On the other hand, systems under *centralized control* achieve globally optimal strategies, and thus achieve *social optimum*. In a centralized or *integrated* system, all the agents operate as an integrated system, i.e., as a single agent interacting with the exterior markets. In our model, the centralized agent has n supply sources with costs for each $\mathrm{s} = (s_1, \ldots, s_n)$ and faces market with demand D, price $p(D)$. A centralized system maximizes the overall profit as well as the total quantity delivered to the consumers.

It is often desirable to give incentives in order to make the decentralized system as efficient as the centralized one, while allowing the agents to have local control. To facilitate this aim, various types of contracts can be suggested. Later, by introducing appropriate contracts, we will show that the agents' equilibrium strategies in a competitive environment can become equivalent to the strategies in a centralized setting. We will also discuss how the centralized and decentralized solutions compare in our models.

In this thesis, the model we study can be described as a two-echelon system: the suppliers constitute the first echelon while both the retailer and the suppliers comprise the second echelon. We next discuss in more detail the loss of efficiency in two-echelon systems.

Generally, notice that there could be several sources of inefficiency in a two-echelon system. These include random shocks or inputs (stochastic consumers' demand in our model), incomplete information (not present in our model since we assume public information), and limited resources (the presence of capacity in our model). However, the major source of inefficiency lies within the sequential structure of the system itself. In a decentralized system where the agents are able to influence the system's output, an effect of *double marginalization* is often present, i.e., the price of the output is higher than the price of the inputs because of two successive mark-ups (marginalizations). On the other hand, in perfectly competitive systems, there is no marginalization at all, as all agents' prices equal their costs. Finally, in monopolies, there is only one mark-up. As a result, competitive systems, i.e., systems under decentralized control, are characterized by greater losses in efficiency when compared to systems under centralized control or to monopolies.

In our model we consider a decentralized system. When the retailer can influence the consumers' price, double marginalization manifests itself through both the suppliers' and the retailer's prices being higher than their respective costs. The total mark-up in the decentralized system is higher than the mark-up in a centralized system. When the prices at the retailer are fixed, a similar effect takes place, as the total amount of the product retailer offers to consumers is smaller than he/she would have offered in a centralized system. That is, the total profit and the total output of a decentralized two-echelon system is not larger than those of a centralized system.

In what follows we present the problem each of the agents faces over the whole time horizon $[0, T]$ in the decentralized environment. We then present the formulation of the centralized problem. We assume that the time horizon is small enough so that discount rates and changes in suppliers parameters' can be omitted.

3.2.2 Decentralized model

The sequence of events

We assume that at each time instance t, the following events occur (see also the illustration in Figure 3-2). First suppliers announce their prices $\mathbf{w}(t) = (w_1(t), \ldots, w_n(t))$; then the retailer responds with rates of order $\mathbf{q}(t) = (q_1(t), \ldots, q_n(t))$ and decides which base stock level $S(t)$ to maintain. Finally, the consumers' demand rate $D(t)$ at the retailer is realized and the profits and losses are assessed by all the agents. The interaction among agents, i.e., the suppliers and retailer, can be modelled as a non-cooperative game. We will describe this game in three ways. First, one can view it as a recurring oligopolistic competition, in which at each time instance, the retailer announces his strategy after all suppliers announced theirs (simultaneously). This game has two stages, and in such interpretation, the suppliers are called *Stackelberg leaders* and the retailer, *a Stackelberg follower*[2]. A second way to view the game is as a competition among only the suppliers. The demand each supplier faces is determined through the retailer's strategy and the consumers' demand. Finally, the third way one can interpret the game is as follows: suppliers announce their prices in the order of decreasing costs and then the retailer follows with an order.

The retailer's problem

We consider the retailer's problem, with suppliers' prices $\mathbf{w}(t)$ as an input. At time instance t, the retailer must decide on the base stock level for the next period $S(t)$ and the rates at which the product is ordered from the suppliers, $\mathbf{q}(t)$, $0 \leq q_i(t) \leq K_i$. We assume that the demand rate is distributed with cdf $F(t)$, $D(t) \sim F(t)$, and $f(t)$ is the corresponding pdf. The price at the consumers' level depends on the demand and is given by function $p(D)$. We assume that $p(D)$ is non-increasing in D. The retailer maintains an inventory $I(t)$. The retailer's revenue is $R(I + q - S, D) = p(D) \min(D, (I + q - S)^+)$. We will denote the instantaneous profit as

$$\pi_r(\mathbf{q}, S, t|\mathbf{w}, I) = R(I+q-S, D) - \mathbf{w}'\mathbf{q} - hS - h(I+q-D-S)^+ - b(I+q-D-S)^-. \quad (3.1)$$

The retailer's problem over time interval $[0, T]$ can be described by the following continuous time dynamic program (DP):

$$\max_{\mathbf{q}, S} E_D \int_0^T \pi_r(\mathbf{q}, S, t|\mathbf{w}, I)dt. \quad (3.2)$$

In this DP,

(i) the state variable is inventory $I(t)$ and the state transition equation is $\dot{I}(t) =$

[2]A *Stackelberg equilibrium* is such that the retailer does not have an incentive to deviate at stage 2 of the problem given that the suppliers play their equilibrium policies, and suppliers do not have incentives to deviate at the first stage of problem.

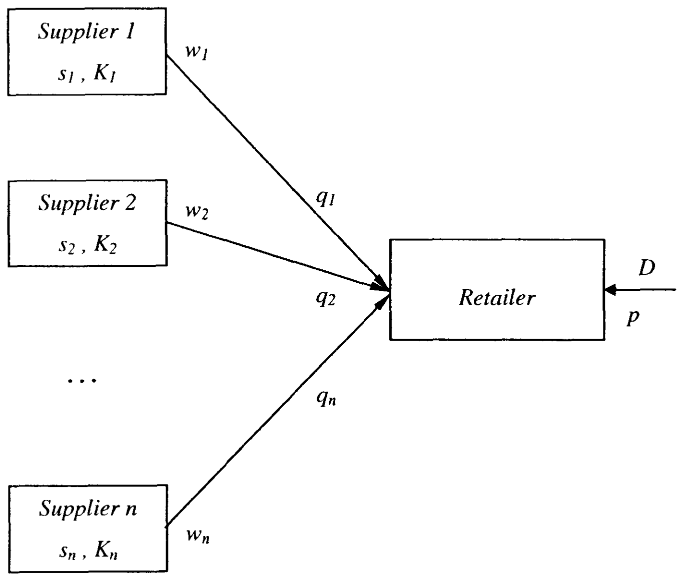

Figure 3-1: Model: n suppliers and a single retailer

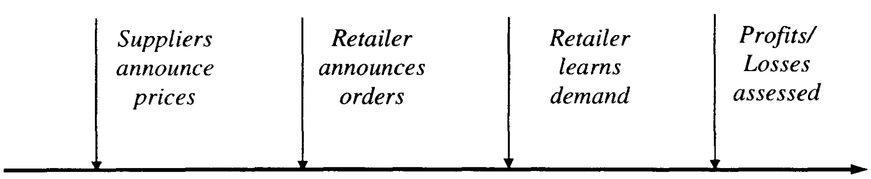

Figure 3-2: Sequence of events

$q(t) - D(t)$ with initial inventory $I(0) = I_0$;

(ii) the control variables are the rates $\mathbf{q}$ at which the product is ordered from the suppliers and the base stock level S, while the control space restriction are $0 \leq q_i(t) \leq K_i$, $0 \leq S(t) \leq I(t)$;

(iii) the randomness comes from the demand, $D(t) \sim F(t)$.

Since our primary goal in this research is to gain an insight into the structure of equilibrium policies, in order to make the computations easier, in the remainder of this thesis we will analyze this DP model through a deterministic fluid approximation. In particular, we will approximate demand rate $D(t)$ through its expected value $E(D(t))$, which we will denote as ED(t) for brevity. Then the retailer will solve the following problem

$$\max_{\mathbf{q},S} \int_0^T \pi_r(\mathbf{q}, S, t|\mathbf{w}, I)dt. \tag{3.3}$$

$$\text{s. t. } \dot{I} = q - ED,\ 0 \leq \mathbf{q} \leq \mathbf{K},\ I \geq S(t) \geq 0,\ I(0) = I_0. \tag{3.4}$$

Alternatively, in order to obtain a closed form solution to the model, we could consider the case of a demand process following a Brownian motion. In this set-up, an appropriate equilibrium can be defined and when the retailer's profit function is quadratic, a closed-form solution can be found (more information on this type of models can be found in [20]).

We assume that when deciding on the orders from equally priced suppliers, the retailer uses a proportional allocation rule, that is, he/she distributes the order between equally-priced suppliers so that the same proportion of their capacities is utilized.

Assumption 13 *The retailer uses a proportional allocation rule.*

This is not a very restrictive assumption, as the model can be analyzed similarly under other allocation rules.

Remark. The proportional allocation rule can have the following interpretation. When several suppliers have the same price w: we could treat them as one supplier with a pooled capacity and then divide the order allocated for that artificial supplier among the pooled suppliers in proportion to their capacity.

Moreover, we assume that all the agents have the same information about technologies, forecasts and strategies of each of the system's agents; for example, each supplier knows the exact costs and capacities of the other suppliers.

Assumption 14 *All information is public.*

Incorporating asymmetric information would enrich the model, however would also change the focus of the research and add a level of difficulty that is beyond the scope of this thesis. Moreover, in mature markets, such as the ones our model represents, we can assume that there is no entry/exit possibilities or technological changes and that the information has dissipated among the agents, i.e. all information is public.

We do not incorporate lead times in our model. This suggests that it does not make sense for the retailer to maintain any positive base stock level, since doing so would only increase inventory costs. Therefore, in the remainder of this discussion, we will omit $S(t)$ from the model.

Assumption 15 *There are no lead times between the suppliers and retailer.*

This is again a simplifying assumption, and the model can generally be extended to include lead times.

We conclude the description of the retailer's problem by the following reformulation, that will be useful in the consequent analysis. The retailer's problem can be interpreted as the competition of $n+1$ suppliers, where the first n suppliers are the actual suppliers as before and the $(n+1)$st supplier is an artificial supplier with capacity $K_{n+1} = ED(t)$ and no costs, $s_{n+1} = 0$. Thus the order from the last supplier represents 'lost demand' (or backorder if such were permitted by the model). Then at each time instance, the holding costs are $h(I + q - ED)$, the costs of lost sales are bq_{n+1}, and, assuming that $w_{n+1} = 0$, the instantaneous profit function can be rewritten as

$$\pi_r(\mathbf{q}, t|\mathbf{w}, I) = R(ED, ED - q_{n+1}) - (\mathbf{w} + h\mathbf{e})\mathbf{q} - bq_{n+1} - hI + hED, \qquad (3.5)$$

and constraints are $\dot{I} = q - ED$, $I \geq 0$, $\mathbf{q} \in [\mathbf{0}, \mathbf{K}]$, and $q + I \geq ED$. The last constraint guarantees that $(n+1)$st supplier correctly represents unsatisfied demand and that inventory stays non-negative. We will also use the following notation: $r(x,t) = R(ED, ED - x, t)$. Hence, $\pi_r(\mathbf{q}|\mathbf{w}, I) = r(q_{n+1}) - (\mathbf{w} + h\mathbf{e})\mathbf{q} - bq_{n+1} - hI + hED$.

A supplier's problem

We next illustrate supplier i's problem. Supplier i has to choose a price level $w_i(t)$ to maximize his/her profit, given the presence of competing suppliers:

$$\max_{w_i} \pi_i(w_i|\mathbf{w}_{-i}) = \int_{t=0}^{T} (w_i - s_i) q_i(w_i, \mathbf{w}_{-i}) dt \qquad (3.6)$$

$$\text{subject to} \quad q_i w_i \leq q_i w_j, \qquad (3.7)$$

$$w_i \geq 0, \qquad (3.8)$$

where $0 \leq q_i \leq K_i$ is determined from the retailer's problem and depends on the state variable I as well as prices $\mathbf{w}$. Constraint (3.7) in the formulation accounts for the competition among the suppliers, i.e., if there is an order from supplier i, with $q_i > 0$, then, at equilibrium, his/her price is not greater than that of the competing suppliers: $w_i \leq w_j$, $\forall j$. Notice that the randomness, that comes into the system with the consumers' demand, is implicitly present in a supplier's problem in function $q_i(w_i, \mathbf{w}_{-i})$. We assume that, at equilibrium, suppliers that do not obtain an order from the retailer price at their costs. This behavioral assumption ensures that the inactive suppliers (i.e., the suppliers without orders in an equilibrium solution) have a greater chance to obtain an order if there is a fluctuation in the active suppliers' offers. Notice that if the equilibrium prices are the result of *a descending auction* being held at each time instance t (i.e., suppliers lower their bids until an equilibrium is reached), then equilibrium prices of active suppliers would be bounded by the costs of inactive suppliers.

Assumption 16 *Suppliers with no orders announce prices equal to their costs.*

Alternatively, supplier i's problem can be modelled with a linear objective depending not just on other suppliers' prices $\mathbf{w}_{-i}$ but also on the retailer's policies $\mathbf{q}$ (see Appendix C). However, in this case, the feasible set is not convex.

The suppliers' problems are complicated due to the presence of competition as well as capacities. Because of these constraints, the model does not necessarily have a pure strategy equilibrium (see Example 1 below).

3.2.3 Centralized problem

The centralized problem finds a solution that would maximize total profit of the integrated system, i.e., it is defined as follows:

$$\max_{q(t)\in[\mathbf{0},\mathbf{K}]} \pi^c(\mathbf{q}(t)|I_0) \quad = \quad E_D \int_0^T (r(q_{n+1}) - \mathbf{s}'\mathbf{q} - h(I+q-D) - bq_{n+1})\, dt, \quad (3.9)$$

where inventory $I \geq 0$, $I(0) = I_0$ is the state variable that follows the transition equation $\dot{I} = q - D$ and $\mathbf{q}$ is the control variable (with $n+1$st supplier denoting "lost sales"). The randomness comes from the demand D. We will refer to the centralized solution as $\mathbf{q}^c$ and the corresponding system's profit as $\pi^c(\mathbf{q}^c|I_0)$. Because this system is under the centralized control, the overall profit extracted from the system is greater than in the decentralized setting. Therefore, potentially, all competing agents can benefit from cooperation or from competition that utilizes some type of contracts.

Notice that the fluid approximation of the problem is as follows:

$$\max_{q(t)\in[\mathbf{0},\mathbf{K}]} \pi^c(\mathbf{q}(t)|I_0) \quad = \quad \int_0^T (r(q_{n+1}) - \mathbf{s}'\mathbf{q} - h(I+q-ED) - bq_{n+1})\, dt, \quad (3.10)$$

with constraints: $I \geq 0$, $I(0) = I_0$, $\dot{I} = q - ED$, and $\mathbf{q} \in [\mathbf{0}, \mathbf{K}]$.

3.2.4 Bilevel programming formulation

The problem over the entire time horizon can be described via a bilevel program as follows:

$$
\begin{aligned}
&\max_{\mathbf{q}} \int_0^T \pi_r(\mathbf{q}, t|\mathbf{w}, I)dt \\
\text{subject to } &w_i = \arg\max_{w_i} \int_0^T \pi_i(w_i, t|\mathbf{q})dt, \\
&\dot{I} = q - ED, \\
&I(0) = I_0, \\
&\mathbf{w} \geq 0,\ \mathbf{0} \leq \mathbf{q} \leq \mathbf{K},
\end{aligned}
$$

where $\pi_i(w_i, t|\mathbf{q}) = (w_i(t) - s_i)q_i(t)$. That is, the retailer orders quantities that maximize his profit subject to the fact that suppliers announce prices that maximize their profits. For more on bilevel programming and related literature see [17].

3.2.5 Continuous time model: analysis

In this section we will establish the existence and analyze structure of the continuous time optimal policies. We first define the notion of equilibrium we use in our model. The model can be viewed as a differential game, where the strategy of a supplier i (player i) is path $w_i(t)$, and the strategy of the retailer (player 0) is path $q(t)$.

Equilibrium in a continuous case

Consider a single period n-player game. Let us denote the strategies of the players by $\mathbf{u}$, such that $u_i \in U_i$ is ith player strategy and U_i is the player's feasible strategy space.

Definition 8 In an n-player the strategies $\mathbf{u}^*$ form a *Nash equilibrium* (NE) if for each player i, u_i^* is the best response to the strategies $\mathbf{u}_{-i}^*$ of the other players.

Notice that when the definition of the space U_i depends on the value of the $\mathbf{u}_{-i}$, an equilibrium in the definition is sometimes referred to as a generalized NE (GNE, see [50]).

In a dynamic setting, consider a truncated dynamic game that starts at time t. Such a subgame depends on the history up to time t, H_t, which is represented in our model by the inventory at time t. Suppose that $\mathbf{u}^*$ is a NE in the original game and $\mathbf{u}_t^*(H_t)$ is its restriction on to the subgame starting at time t.

Definition 9 A NE strategy profile is *subgame perfect*, if for every H_t, the restriction of the profile on the subgame starting at time t with history H_t is a NE of the truncated game as well.

Notice that subgame-perfectness ensures that game's equilibria are "rational" by eliminating incredible threats. Markov perfect equilibria are a further sharpening of the equilibria concept.

Definition 10 A *Markov-perfect NE* (MPNE) is a subgame perfect NE in which each strategy profile depends on the state variables at time t only.

(For more discussion of these types of equilibria see [20] and [24].)

In our model since all the history of the game so far is encapsulated within one variable: $I(t)$, every subgame perfect equilibrium is also a MPNE.

These definitions apply to both continuous and discrete games; however, theoretical results on continuous and, generally, on non-finite games are harder to obtain.

In the next section, we examine in more detail the solutions of each problem that we introduced so far in a continuous time setting. But before proceeding, we make the following note on the equilibrium policies. Suppose that the suppliers announce their equilibrium prices $\mathbf{w}^*$ and that the retailer's equilibrium orders are $\mathbf{q}^*$. We can equivalently characterize the equilibrium by a pair of vectors $(\mathbf{w}^*, \mathbf{q}^*)$ or a pair of scalars (w^*, q^*), where w^* is the price announced by the active suppliers (when the policy is pure, such price is well defined) and q^* is the total amount that the retailer orders from all the suppliers. Notice that $\mathbf{w}_i^* = w^*$, for i such that $q^* > K^{i-1}$ and $\mathbf{w}_i^* = s_i$ otherwise. Henceforth, the references to $\mathbf{w}^*$ and w^*, $\mathbf{q}^*$ and q^* will be used interchangeably.

The retailer's problem

We consider a solution on a small interval $[\tau, \tau+\delta)$ and assume that w^* is continuous on this interval. In order to solve the retailer's problem, we take a similar approach to Bertsimas and Paschalidis [5]: we use (a) Pontryagin's maximum principle, (b) a small interval approximation, and (c) continuity assumptions. Nevertheless, the context and application we consider are different.

Lemma 5 *Suppose that* $\mathbf{w}^*$ *is an equilibrium policy for the suppliers. Suppose that on* $[\tau, \tau+\delta)$ *this policy is continuous, then there exists a measurable retailer's equilibrium policy* $\mathbf{q}^*(\mathbf{w}^*)$.

Proof. This result follows from the existence result by Filippov-Cesari (see [29]), since both the state and control variables in our problem are bounded. ■

We next take on the retailer's problem (3.1) using Pontryagin's Maximum Principle.

We construct the Hamiltonian for the retailer's problem by releasing the state transition equation:

$$H(I(t), \mathbf{q}(t), \lambda(t)) = \pi_r(\mathbf{q}|\mathbf{w}(t), I(t), ED(t)) + \lambda(t)(q(t) - ED(t)).$$

Next we construct the Lagrangian for the problem by releasing the capacity, non-negativity and supply constraints:

$$L(I, \mathbf{q}, \lambda, \nu^1, \nu^2, \nu, \zeta) = H(I, \mathbf{q}, \lambda) + \nu^1 \mathbf{q} + \nu^2(\mathbf{K} - \mathbf{q}) + \nu(q + I - ED) + \zeta(q - ED),$$

where the last term corresponds to the state transition equation.

The necessary conditions for a trajectory I^* to be optimal (see [52]) are:

$$\begin{cases} L_I = -\dot{\lambda} + \nu, \; L_\lambda = \dot{I}, \\ \nu^{1'}\mathbf{q} = 0, \; \nu^{2'}(\mathbf{K} - \mathbf{q}) = 0, \; \nu^1, \nu^2 \geq 0, \; 0 \leq \mathbf{q} \leq \mathbf{K}, \\ \nu \geq 0, \; q + I \geq ED, \; \nu(q + I - ED) = 0, \\ \mathbf{q}^* \text{ is a maximizer of } H(I^*, \mathbf{q}, \lambda) \text{ s.t. } 0 \leq \mathbf{q} \leq \mathbf{K}, \; q^*(t) + I^*(t) \geq ED(t), \\ \lambda(\tau + \delta) = 0, \; \text{(transversality condition)} \\ \zeta I = 0, \; I \geq 0, \; \zeta \geq 0, \; \dot{\zeta} \leq 0, \end{cases} \tag{3.11}$$

In this setting,

$$\frac{\partial \pi_r}{\partial q_i} = -\,(w_i + h), \;\; i = 1, \cdots, n; \tag{3.12}$$

$$\frac{\partial \pi_r}{\partial q_i} = -\,(h + b) + r'(q_i), \;\; i = n + 1; \tag{3.13}$$

$$\frac{\partial \pi_r}{\partial I} = -\,h. \tag{3.14}$$

The following lemma summarizes the properties of the retailer's instantaneous profit function.

Lemma 6 *The retailer's instantaneous profit is a strictly decreasing linear function in the state variable I and control variables q_i, $i = 1, \cdots, n$. Moreover, as long as the marginal revenue associated with a lost sale is smaller than $b + h$, the profit is also strictly decreasing in q_{n+1}. The profit is concave in q_{n+1} when the revenue is concave in this variable.*

Proof. This follows from the expressions for the first derivatives as shown in (3.12)-(3.14). ■

When $I(t) > 0$ then $\zeta(t) = 0$. Therefore, since $\zeta(t) \geq 0$, $\dot{\zeta}(t) \leq 0$ and ζ is continuous, $\zeta(t) = 0$ when $I(t) = 0$. That is, $\zeta(t) = 0$.

Therefore, the Maximum Principle conditions (3.11) imply that

$$\begin{cases} \dot{\lambda} = -\frac{\partial L}{\partial I} = -\frac{\partial \pi_r}{\partial I} + \nu = -h + \nu, \\ 0 = \frac{\partial H(I^*, q, \lambda)}{\partial q_i} + \lambda + \nu_i^1 - \nu_i^2 + \nu = \frac{\partial \pi_r}{\partial q_i} + \lambda + \nu_i^1 - \nu_i^2 + \nu \\ \qquad = \begin{cases} -w_i - h + \lambda + \nu_i^1 - \nu_i^2 + \nu, & \text{for } i = 1, \cdots, n, \\ r'(q_i) - h - b + \lambda + \nu_i^1 - \nu_i^2 + \nu, & \text{for } i = n+1, \end{cases} \\ \nu_i^1, \nu_i^1 \geq 0,\ \nu_i^1 q_i = 0,\ \nu_i^2 (q_i - K_i) = 0,\ 0 \leq q_i \leq K_i, \\ \nu \geq 0,\ \nu(q + I - ED) = 0,\ q + I \geq ED. \end{cases} \tag{3.15}$$

Since we assumed that the equilibrium policy is pure, as is shown in Theorem 8 below, for all active [real] suppliers it must hold that $q_i = K_i$. Therefore, the structure of the retailer's solution implies that, if multiplier $\lambda(t)$, that corresponds to some pure equilibrium, is known, the retailer can compute the quantities to be ordered using a greedy rule as follows:

A greedy allocation rule for the retailer

Step 1 Sort suppliers in order of increasing prices $w_i(t)$. Let $q^*(t) = 0$.

Step 2 Starting from the lowest priced supplier, $i = 1$:

(a) If (i) $I^* + K^i \leq ED$ and $w_i < b - r'(ED - K^i - I^*)$ or if (ii) $I^* + K^i > ED$, $w_i < b - r'(ED)$, then award supplier i with $q_i^*(t) = K_i$ and let $q^*(t) = K^i$.

Repeat Step 2 for the next supplier.

(b) Else, award supplier $j \geq i$ with $q_j^*(t) = 0$. The artificial supplier gets $q_{n+1}^* = (ED - I^* - q^*)^+$.

Stop.

Hence, we can see that the total amount ordered from real suppliers

$$\sum_1^n q_i^*(t) = \max_i K^i 1_{w_i < b - R'(\min(ED, I + K^i))}$$

and $q_{n+1}^* = (ED - I^* - q^*)^+$. Moreover, $\nu = (-R'(ED, q^*) + h + b - \lambda) 1_{I^* + q^* < ED}$.

When $I^* + q^* < ED$, $\lambda = -h(t - \tau - \sigma)$, while when $I^* + q^* > ED$, the following differential equation holds for λ: $\dot{\lambda} + \lambda - b + r'(ED - q^*) = 0$. Notice, that q^*, as described above, essentially is some function of $\lambda(t)$ and $w^*(t)$. We can rewrite the equation for λ as $\dot{\lambda} + \lambda - b + r(w^*, \lambda) = 0$, and $\lambda(t) = l(w^*(t))$ would be a solution of this equation if such exists, where $l(\cdot)$ is some functional of $w^*(t)$.

Next notice that $\dot{I} = K^i 1_{b-r'(ED-K^i-I)>w_i} - ED$. In this expression i depends on t and $ED(t)$; if these parameters were constant, we would obtain that I is a linear function of t.

We conclude this analysis with the following results characterizing the solution of the retailer's problem:

Proposition 6 *Let $\lambda(t)$ be an optimal multiplier associated with the state transition equation; $\mathbf{w}^*$, an equilibrium suppliers' policy and I^*, an inventory at instant t. The optimal order quantities for the retailer's problem should satisfy the following conditions:*

$$q_i^* = \begin{cases} K_i, & \text{if } w_i^* + \lambda + h - r'(ED - \min(I^* + K^i, ED)) \geq 0, \\ 0, & \text{if } w_i^* + \lambda + h - r'(ED - \min(I^* + K^i, ED)) < 0. \end{cases} \tag{3.16}$$

Remark. If the total amount ordered by the retailer is limited by some $\tilde{K} > 0$, $q \leq \tilde{K}$, then the controls of the modified model q_i^{constr}, i.e., the quantities ordered, become $q_i^{constr} = \min\{q_i^{unconstr}, \left(\tilde{K} - \sum_{j:w_j<w_i} K_i\right)^+\}$, where the quantities $q_i^{unconstr}$ are found in (3.16).

Proposition 7 *Suppose that on interval $[\tau, \tau + \delta)$, there exists an integrable equilibrium policy $w^*(t)$ (i.e., a price for the active suppliers). Then the adjacent multiplier $\lambda(t)$ can be expressed as a functional of $w^*(t)$ on interval $[\tau, \tau + \delta)$. Moreover, $I(t)$ and $\mathbf{q}(t)$ are functionals of $w^*(t)$ and an initial inventory at time τ, for $\forall t \in [\tau, \tau + \delta)$.*

Proof. This follows from the discussion above. ■

The continuous time case we studied above provides us with an initial approximation of the system's behavior in the context of a mature market with frequently recurring oligopolistic competition. In the next subsection we conclude the description of the continuous time case presenting the suppliers' strategies and discussing the conditions for the existence of a pure equilibrium in more detail.

A supplier's problem

When a pure strategy equilibrium is considered, at any time instance, supplier i's strategy is bounded from above by some $w_i^*(t)$, where $w_i^*(t)$ depends on the strategies of the other suppliers. For active suppliers, this bound equals to the equilibrium policy $w^*(t)$ that we considered in the retailer's problem. For inactive suppliers, Assumption 16 implies that $w_i^*(t) = s_i$. Hence, $w_i^*(t) = \max(w^*(t), s_i)$. Moreover, the upper bound in the active suppliers' problems is the same as the borderline price w^* that separates prices of active and inactive suppliers as in (3.16). Thus, as we observed in the previous subsection, at an equilibrium we can express the upper bound in the active the supplier's problem w^* as a functional of multiplier λ in the retailer's problem.

Observe that each supplier's problem is time-separable as long as the value of the inventory from the retailer's problem is known. Therefore, at any instant t, supplier i's problem can be formulated as

$$\max_{w_i}\{(w_i - s_i)q_i(w_i, \mathbf{w}_{-i}) \mid w_i \geq 0, \} \tag{3.17}$$

where $q_i(\mathbf{w})$ is determined through solving the retailer's problem. This is a non-concave, non-continuous optimization problem over a closed convex set. It is non-continuous since $q_i(\mathbf{w})$ is not a continuous function of $\mathbf{w}$. Moreover, the function $\pi_i(w_i|\mathbf{w}_{-i})$ is not concave or continuous, or even upper-semicontinuous (i.e., as $w_i^n \to w_i$, it does not necessarily hold that $\limsup_n \pi_i(w_i^n|\mathbf{w}_{-i}) = \pi_i(w_i, |\mathbf{w}_{-i})$). In consequence, the standard equilibrium results for continuous games ([24]) do not apply. Furthermore, the theory of discontinuous games (Dasgupta and Maskin [16], [24]) does not apply either. Therefore, a pure strategy equilibrium can exist in our model under additional assumptions. In the next subsection, we study some of these assumptions guaranteeing the existence of a pure strategy equilibrium.

3.2.6 Equilibrium in pure strategies

Introductory example

We first introduce an example (Example 1) of static competition between two suppliers in order to illustrate some of the issues, such as the existence of pure strategy equilibrium as well as the computation of the equilibrium policies, on the case of static (single period) competition among two suppliers and a single retailer. Thereafter, we will present a multi-period version of the model we considered in Example 1. We will also present conditions for existence of a pure strategy equilibrium in both discrete and continuous time settings.

Example 1 The retailer will determine his/her total order q and $p(q) = a - bq$ based on maximizing his/her profit function $\pi_r(\mathbf{q}) = p(q)q - \mathbf{w}'\mathbf{q}$, where $q = q_1 + q_2$, $q_i \in [0, K_i]$ is an order from supplier i and w_i is a price announced by supplier i.

Suppose that the strategy of supplier two is w_2, then the best response policy from supplier one computed by solving the profit-maximization problem: $\max_{w_1} \pi_1(\mathbf{w}) = (w_1 - s_1)q_1(\mathbf{w})$, is the following:

$$w_1(w_2) = \begin{cases} s_1 & q_1(\mathbf{w}) = 0 & \text{if } s_1 > w_2, \\ w_1^m = \frac{a+s_1}{2}, & q_1(\mathbf{w}) = \frac{a-w_1}{2b}, & \text{if } \frac{a+s_1}{2} \leq w_2, \\ w = \min(a - 2bK^2, \frac{a+s_1}{2} - bK_2), & q_1(\mathbf{w}) = \frac{a-w_1}{2b} - K_2, & \text{if } w \geq w_2, \\ w_2 - \varepsilon, & q_1(\mathbf{w}) = K_1, & \text{otherwise,} \end{cases}$$

where ε is a very small positive number, such that supplier one is able to undermine supplier two's policy, and w_1^m is supplier one's monopoly price. We will generally omit ϵ when we characterize the equilibrium.

The profit for supplier i can be calculated from the best response strategy expression we described above, that is $\max \pi_i(\mathbf{w}) = (w_i - s_i)q_i(\mathbf{w})$ subject to $q_i(\mathbf{w}) \in [0, K_i]$.

The existence of a pure strategy equilibrium $\mathbf{w}^*$ implies that $\frac{\partial \pi_i(\mathbf{w}^*)}{\partial w_i} \leq 0$, where

$$\frac{\partial \pi_1(\mathbf{w})}{\partial w_1} = \begin{cases} 0, & \text{if } w_2 < s_1, \\ \frac{a+s_1-2w_1}{2b}, & \text{if } \frac{a+s_1}{2} \leq w_2, \\ \frac{a+s_1-2w_1}{2b} - K_2, & \text{if } \frac{a+s_1}{2} - bK_2 \geq w_2 \geq s_1, \\ K_1, & \text{otherwise,} \end{cases}$$

and $\frac{\partial \pi_2(\mathbf{w})}{\partial w_2}$ is defined in a similar way.

Let us assume that $a \geq s_i$ and $s_1 \neq s_2$. Under these assumptions, a pure strategy equilibrium exists in two cases:

(1) when one of the suppliers can enforce a monopoly: i.e., $q_i = \min(q_m, K_i)$, $q_{-i} = 0$ or

(2) when both suppliers are fully engaged in the market, i.e., $q_i = K_i$.

The first condition would hold if for some i: $\frac{a+s_i}{2} < s_{-i}$. The second condition would hold when $a - 2bK^2 \geq \max(\frac{a+s_i}{2} - bK_{-i}, s_i)$.

Continuous case

Before characterizing equilibrium policies for the continuous competition, we note that if supplier i is active in an equilibrium solution (i.e., $q_i > 0$), then so is every supplier j with marginal cost $s_j \leq s_i$. Let us denote the retailer's objective starting at time t as $\Pi^t(\mathbf{q}^t, I^t) = \pi_r^t(\mathbf{q}^t|\mathbf{w}^t, I^t) + DP^{t+1}(\mathbf{w}, I^t + q^t - ED^t)$.

Theorem 8 *Suppose Assumptions 13 - 16 hold. The following conditions are necessary for a pure strategy* $(\mathbf{w}^*, \mathbf{q}^*)$ *to be an MPNE. Moreover, condition (1) is sufficient as well.*

(1) For all active suppliers: $\frac{\partial \Pi_i(w_i^*|\mathbf{w}^*_{-i})}{\partial w_i} \leq 0$ *and* $w_i \leq \min(s_{k+1}, p(ED))$ *where* k *is the number of active suppliers (we assume that* $s_{n+1} = \infty$*),*

(2) If the number of active suppliers $k > 1$ *then* $q^* = K^k$*.*

Proof. *Necessity.* From the definition of an MPNE and since suppliers problem is time separable (in the presence of retailer's strategy (q^{*t}), if follows that w^{*t}_i should be a maximizer of supplier i's profit function, $\pi_i(w_i|\mathbf{w}^{*t}_{-i}, q^{*t})$. Hence any pure equilibrium strategy would satisfy (1).

We will show by contradiction that condition (2) must hold. Suppose that at equilibrium there are more than two active suppliers and that one of them, supplier i, gets an order $q_i^*(t) \in (0, K_i)$. Then, since $w^*(t) > s_i$ (otherwise the order from supplier would be $q_i^*(t) = 0$) and since there is at least one other supplier from whom the retailer orders, it follows that by lowering his/her price a little, supplier i can receive an order $q_i(t) > q_i^*$ and, thereby, increase his/her overall profit - this is a contradiction.

Sufficiency. Suppose that (1) and (2) hold, but at some time t, the first k suppliers do not constitute an equilibrium at $\mathbf{w}^*$ as described in the statement of the theorem. Then it must be that supplier $k+1$ can enter the market with some positive profit. Then $\exists w > w^{*t}$ such that $w > s_{k+1}$ and $q^t_{k+1} > 0$ at some different equilibrium policy. That would imply, however, that each supplier $i \leq k$ can be improve his/her profit by increasing w^{*t}_i. This contradicts condition (1). ■

Moreover, recall that under Assumption 16, every inactive supplier i sets the price as $w_i^* = s_i$ at equilibrium.

Example 2 We will illustrate here that the condition (1) can be satisfied by the suppliers. Suppose that $p(x) = a - bx$ is the price that a consumer pays per unit where $x \leq D$ is the amount of product that is available at the retailer. Then the retailer's revenue function at time t is quadratic as in $r(q^t_{n+1}) = (a - b(ED^t - q^t_{n+1}))(ED^t - q^t_{n+1})$. Then the profit of an active supplier i is as in in Figure 3-3. Notice that this is a quasiconcave function, and that the set of supergradients at point $w = w^{*t}$ consists of only those sloping down. Thus condition (1) is satisfied.

At this point we reformulate the problem of finding equilibrium can policies as a continuous time $DVIP$ with problem function $F(\mathbf{w}, \mathbf{q}) = (-\frac{\partial \pi_r(\mathbf{q}|\mathbf{w}, I)}{\partial \mathbf{q}}, -\mathbf{q})$.

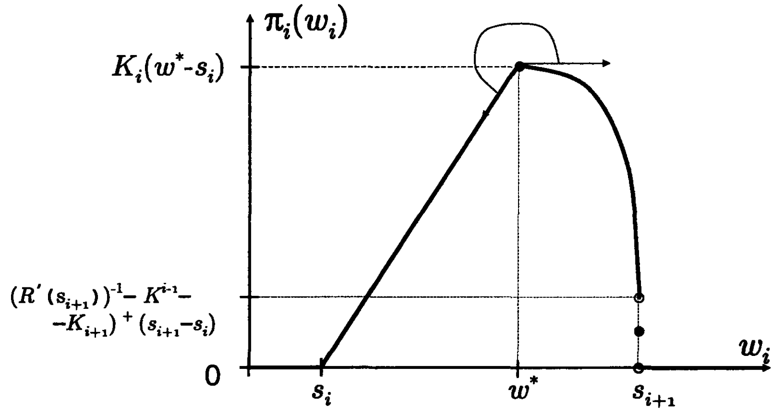

Figure 3-3: Profit function of an active supplier

Proposition 8 *Suppose $z^* = (\mathbf{w}^*, \mathbf{q}^*)$ is a pure strategy MPNE, then it satisfies*

$$\int_0^T F(z^*)'(z - z^*)dt \geq 0 \textit{ where } z_q \in [\mathbf{0}, \mathbf{K}],\; z_w \in W(z_q),\; I(0) = I_0,\; \dot{I} = q - ED. \tag{3.18}$$

Proof. It follows immediately from (3.16). ∎

3.3 Discrete time model

3.3.1 Formulation

In this subsection, we present a discrete time model. The construction is similar to that in the continuous time model in the previous section and we omit it for the sake of brevity. The reason for presenting a discrete time model is that it allows us to consider efficient solution methods but also to get insight through studying additional properties.

In particular, in the deterministic fluid model approximation, the state transition equation becomes $I^{t+1} = I^t + q^t - ED^t$, with initial condition $I^1 = I_0$. The retailer's profit per period t becomes $\pi_r^t(\mathbf{q}^t|\mathbf{w}^t, I^t, ED^t) = r(q_{n+1}^t) - (\mathbf{w}^t + h\mathbf{e})\mathbf{q}^t - bq_{n+1}^t - hI^t + hED^t$, where $r(q_{n+1}^t)$ is the revenue function given demand ED^t at time t. The retailer's objective is

$$\max_{\substack{0 \le q_i^t \le K_i, \\ I^t + q^t \ge ED^t}} \left\{ \sum_1^T \pi_r^t(\mathbf{q}^t | \mathbf{w}^t, I^t, ED^t) \,|\, I^{t+1} = I^t + q^t - ED^t, \; I^1 = I_0 \right\}. \tag{3.19}$$

Notice that given initial inventory I^t at time t and the suppliers' prices $\mathbf{w}$, the retailer's optimal profit-to-go, $DP(\mathbf{w}, I^t)$, satisfies the following recursion:

$$DP^t(\mathbf{w}, I^t) = \max_{\substack{0 \le q_i^t \le K_i, \\ I^t + q^t \ge ED^t}} \left\{ \pi_r^t(I^t, \mathbf{q}^t, ED^t) + DP^{t+1}(\mathbf{w}, I^t + q^t - ED^t) \right\}.$$

Supplier i's problem is also defined similarly to the continuous time case:

$$\max_{w_i^t \ge 0} \left\{ \sum_{t=1}^T (w_i^t - s_i) q_i^t \mid w_i^t \le \max(\min \mathbf{w}_{-i}(q^t), s_i) \right\}. \tag{3.20}$$

In what follows, we denote the retailer's problem starting at time t, with initial inventory I^t as $DP^t(\mathbf{w}^t, I^t)$, where $\mathbf{w}^t$ is the vector of prices starting at period t. We use the same notion for the value of the retailer's problem when it is unambiguous. We will also use the notion of an MPNE to characterize equilibria in this retailer-suppliers game.

3.3.2 *DVIP* reformulation

We next formulate the multi-period problem as a dynamic variational inequality problem ($DVIP$). We define a mapping $F : [\mathbf{0}, \mathbf{K}] \times W \times \mathbb{R}^+ \longrightarrow \mathbb{R}^{2n+1}$:

$$F(\mathbf{q}, \mathbf{w}, I) = \left(-\frac{d\pi_r(\mathbf{q}|\mathbf{w}, I)}{d\mathbf{q}}, \mathbf{q}, h \right).$$

Let $W_i(\mathbf{q}) = \{w \mid w \ge s_i, \; (w_i - w_j) q_i(\mathbf{w}) \le 0, \forall j\}$. Notice that function $\mathbf{q}(\mathbf{w})$ and, hence, the structure of the feasible set can be derived from the optimality conditions for the retailer. Notice that the problem function $-F$ has the same monotonicity properties as R', i.e., F is monotone as long as R is concave. Even when the $DVIP$

function is monotone, since the feasible space is not a convex set, the theory for monotone VIPs does not apply readily to our problem. However, the following equivalence result holds.

Proposition 9 *$x^{t*} = (\mathbf{q}^{t*}, \mathbf{w}^{t*}, I^{t*})$ is a pure strategy MPNE if and only if it satisfies the following $DVIP$:*

$$\sum_{1}^{T} F^I(x^{t*})'(x^t - x^{t*}) \geq 0,$$

where $\mathbf{q}^t \in [0, \mathbf{K}],\ \mathbf{w}^t \in W^t(q^t),\ I^{t+1} = I^t + q^t - ED^t,\ I^{t+1} \geq 0,\ I^1 = I_0.$

Proof. $DVIP$ $\Rightarrow$ MPNE.

As is well known from dynamic programming theory, the fact that $(\mathbf{q}^*, \mathbf{w}^*, I^*)$ is a solution of the above $DVIP$ implies that $\mathbf{q}^*, \mathbf{w}^*$ are solutions of each period's profit-to-go problems for the retailer or for a supplier. Let $\mathbf{w}^t = \mathbf{w}^{t*}$, $\forall t$, $\mathbf{q}^t = \mathbf{q}^{t*}$, $\forall t < k$, $I^t = I^{t*}$, $\forall t < k$, the above $DVIP$ becomes

$$\sum_{t=k}^{T} \pi_{rq}^t(\mathbf{q}^{t*}|\mathbf{w}^{t*}, I^{t*})'(\mathbf{q}^t - \mathbf{q}^{t*}) + \pi_{ri}^t(\mathbf{q}^{t*}|\mathbf{w}^{t*}, I^{t*})(I^t - I^{t*}) \leq 0,$$

where π_{rq} is the first derivative of the retailers profit function with respect to $\mathbf{q}$ and π_{ri}, with respect to I. This inequality implies that there does not exist a direction of ascent for the retailer's problem $DP^k(\mathbf{w}^{k*}, I^{k*})$ at point $\mathbf{q}^{t*}, I^{t*}, t \geq k$, i.e., this point is an optimal solution for the problem $DP^k(\mathbf{w}^{k*}, I^{k*})$. Similarly, one can show that for each supplier i, w_i^{t*}, $t = 1, \cdots, T$ is an optimal response policy. Hence $(\mathbf{q}^*, \mathbf{w}^*, I^*)$ is a subgame perfect equilibrium of the underlying dynamic game.

MPNE $\Rightarrow$ $DVIP$

This follows directly from the definition of MPNE. Since $(\mathbf{q}^*, \mathbf{w}^*)$ is a set of the best response strategies, for the retailer's problem it must hold that $-\frac{d\pi_r(\mathbf{q}^*, \mathbf{w}^*, I^*)}{d\mathbf{q}}(\mathbf{q} - \mathbf{q}^*) + h(I - I^*) \geq 0$, for every feasible $\mathbf{q}$ and I. Similarly it must hold that $\mathbf{q}^{*\prime}(\mathbf{w} - \mathbf{w}^*) \geq 0$ for every $\mathbf{w} \in W(q^*)$. Otherwise, better policies would exist and one of the agents could modify his strategy so as to achieve a greater profit. ■

From the $DVIP$ formulation, several characteristics of the equilibrium policies can be deduced immediately. We summarize them in the following proposition.

Proposition 10 *Suppose* $(\mathbf{w}^*, \mathbf{q}^*)$ *is a pure MPNE. Then*

(1) If $q^{*t}_i > 0$ *then* $w^*_i = \min(\mathbf{w}^{*}_{-i}{}^t, p(ED^t)) = w^{*t}$,

(2) If $w^{*t}_i > w^{*t}$ *then* $w^{*t}_i = s_i$, $q^{*t}_i = 0$,

(3) $w^{*t} \in [s_k, s_{k+1}]$ *when* $q^{*t} \in [K^{k-1}, K^k)$,

(4) $q^{*t}_i = \min(q^{*t} - K^{i-1}, K^i)1_{w^{*t}_i \leq p(ED^t)}$.

Proof.

(1) From the formulation of supplier i's problem, it follows that for every active supplier i, $w^*_i \leq \min(\mathbf{w}^*_{-i}, p(ED))$. Suppose, this inequality is strict. Then some feasible $w > w^*_i$ would guarantee supplier i a strictly greater profit.

(2) Using the result in item (1), we can see that $w^{*t}_i > w^{*t}$ implies that supplier i is inactive, and hence by Assumption 16, $w^{*t}_i = s_i$.

(3) This follows from items (1) and (2) and the fact that the suppliers compete.

(4) This follows from items (1) and (3). ■

It is possible to reformulate the problem as a $DVIP$ with a convex feasible set, however in that case the problem function would not be monotone. (Appendix C).

3.3.3 The retailer's problem properties

We next present the discrete time approximation and analysis of the solution. It is similar to the one in Subsection 3.2.5:

The Hamiltonian: $H(\mathbf{q}^t, I^t, \lambda^t) = \pi^t_r(\mathbf{q}^t, I^t) + \lambda^t(q^t - ED^t)$.

The adjoint equation: $\lambda^t = \lambda^{t+1} + H_{I^t}$, $\lambda^T = 0$.

The optimal control: $\mathbf{q}^{*t} = \arg\max_{0 \leq q_i \leq K_i} H(\mathbf{q}, I^{*t}, \lambda^{t+1})$.

From the adjoint equation, since $H_I = -h$, we derive that $\lambda^t = -h(T-t)$.

Therefore,

$$\mathbf{q} = \arg \max_{\substack{q \geq ED - I, \\ \mathbf{q} \in [0, \mathbf{K}]}} r(q_{n+1}) - bq_{n+1} - (h(T-t+1)\mathbf{e} + \mathbf{w})'\mathbf{q},$$

recall that q_{n+1} is the quantity of unfilled customer's orders. The solution to this problem is

$$q_i = \begin{cases} 0, & \text{if } p(ED) + b < w_i \text{ or } \sum_{w_j < w_i} K_j \geq ED - I, \\ K_i \frac{K_i}{\sum_{w_j = w_i} K_j}, & \text{otherwise (using Assumption 13).} \end{cases} \tag{3.21}$$

Notice that the quantities ordered depend on only on customers' demand and the initial inventory at time t.

3.3.4 Properties of the profit-to-go function

Proposition 11 *The retailer's optimal control policy $\mathbf{q}(I)$ exists for every vector of suppliers' prices.*

Proof. The optimization problem that finds the profit-to-go function $DP^t(\mathbf{w}, I^t)$ is a problem over a non-empty, convex, closed space (the set $\{\mathbf{q}^t \,|\, 0 \leq \mathbf{q} \leq \mathbf{K}\}$). Thus, an optimal solution always exists, i.e., the profit-to-go function is well defined. ■

Even though a solution to the retailer's problem exists, the maximized function $\Pi^t(\mathbf{q}^t, I^t)$ is not necessarily convex. The following proposition proves some properties pertaining to the retailer's problem. We will omit time and suppliers' prices from the notation and will denote $DP^t(\mathbf{w}, I^t)$ as $DP(I)$ and DP^{t+1} as DP^+, where $t < T$.

Proposition 12 *(a) $DP(I)$ is an increasing piecewise linear function of I, (b) $q(I)$ is non-increasing, (c) the slope of $DP(I)$ is non-increasing, (d) $DP(I)$ is continuous on the left, and (e) $\Delta DP(I) = DP(I+1) - DP(I)$ is non-decreasing in I on the domain without the discontinuity points.*

Proof. First observe that for the suppliers' equilibrium prices $w_i(q) \geq w_i(q - \theta)$, where q is the total amount ordered by the retailer during a particular period and

$\theta \geq 0$. For inactive supplier, this relation holds with equality sign ($K^i \geq q$). For the other (active) suppliers, the equilibrium price increases, as the total amount that the retailer orders increases. Therefore, the inequality above will also hold.

We prove the proposition using backward induction, namely we suppose that for DP^+ the statement of the proposition holds (notice that proposition holds for a one-period game).

(a) First, observe that $DP(I)$ is piecewise linear since it is a sum of piecewise linear functions. Suppose that $\hat{q}$ is an optimal solution of $DP(I+1)$, while q^* an optimal solution of $DP(I)$. Then

$$\begin{aligned}
DP(I) &= r(q^*_{n+1}) - bq^*_{n+1} - (h + w(q^*))q^* + DP^+(I + q^* - ED) \\
&= r(q^*_{n+1}) - bq^*_{n+1} - (h + w(q^* - 1))(q^* - 1) \\
&+ DP^+((I+1) + (q^* - 1) - ED) + (w(q^* - 1)(q^* - 1) - w(q^*)q^*) - h \\
&\leq DP(I+1) + (w(q^* - 1)(q^* - 1) - w(q^*)q^*) \leq DP(I+1).
\end{aligned}$$

(b) Similarly to the derivations in the previous item, we can obtain that

$$\begin{aligned}
DP(I+1) &= r(\hat{q}_{n+1}) - b\hat{q}_{n+1} - (h + w(\hat{q}+1))(I + (1+\hat{q}) - ED) \\
&+ DP^+\left((I + (1+\hat{q}) - ED)^+\right) - w(\hat{q})'\hat{q} + w(\hat{q}+1)(\hat{q}+1) \\
&\leq DP(I) - w(\hat{q})'\hat{q} + w(\hat{q}+1)(\hat{q}+1).
\end{aligned}$$

In summary, we obtained the following bounds:

$$DP(I) + w(q^*)q^* - w(q^*-1)(q^*-1) \leq DP(I+1) \leq DP(I) + w(\hat{q}+1)(\hat{q}+1) - w(\hat{q})(\hat{q}).$$

Let us denote by $\Delta w(x)(x)$ a marginal change in the value of $w(x)x$ as x goes from x to $x+1$. Then, we have shown that $\Delta w(\hat{q})(\hat{q}) \geq \Delta w(q^* - 1)(q^* - 1)$. Since $w(x)x$ is increasing in x with an increasing slope and continuous on the left, it follows that $\hat{q} \geq q^* - 1$.

(c) We will next show that the slope of $DP(I)$ is non-increasing. Let $p = p(ED)$, and denote by $slope^*$ the slope of the retailer's profit function $\pi_r(\mathbf{q}|\mathbf{w}, I)+DP^+(I+q-ED)$ at point q^*; by $\widehat{slope}$, at point $\hat{q}$; by $slope_+(x)$, the slope of the retailer's profit function $DP^+(x)$. Then

1. $\hat{q} = q^*-1$, hence $I+q^*-1 \geq ED$ and $slope^* = -h+slope^*_+$, $\widehat{slope} = -h+\widehat{slope}_+$. Therefore, $\widehat{slope} - slope^* \leq 0$ by hypothesis.

2. $\hat{q} = q^*$, $\widehat{slope} - slope^* \leq 0$ by hypothesis.

3. $\hat{q} \geq q^* + 1$. Notice that $\hat{q} \neq q^* + 1$. Suppose, $\hat{q} \geq q^* + 2$. Then if $I + q^* \geq ED$, $\widehat{slope} = -h + slope_+(I + \hat{q} - ED) \geq -h + slope_+(I + q^* - ED)$ by hypothesis. When $I+q^* < ED$ but $I+\hat{q} \geq ED$, we obtain $\widehat{slope} = -h+slope_+(I+\hat{q}-ED) \geq p + slope_+(I + q^* - ED)$. Finally, when $I + \hat{q} < ED$, we also obtain that $\widehat{slope} = p + slope_+(I + \hat{q} - ED) \geq p + slope_+(I + q^* - ED)$ from the initial hypothesis.

(d) $DP(I)$ is continuous on the left, i.e., $\lim_{\sigma \to 0} DP(I - \sigma) = DP(I)$.

Suppose $I + q > ED$. Then for a small enough $\sigma > 0$, it follows that $DP(I - \sigma) - DP(I) \geq \sigma h + DP^+(I + q - ED - \sigma) - DP^+(I + q - ED) \to 0$ as $\sigma \to 0$. This follows from the induction hypothesis.

We also need to consider the case when $I+q \leq ED$. As before, for a small enough $\sigma > 0$, we obtain $DP(I - \sigma) - DP(I) \geq -p\sigma + DP^+(I + q - ED - \sigma) - DP^+(I + q - ED) \to 0$ as $\sigma \to 0$. This follows from the induction hypothesis.

(e) Finally, we will prove that $\Delta DP(I) = DP(I + 1) - DP(I)$ is non-decreasing in I on the domain without discontinuity points. This follows from the items we proved above: the left continuity and $DP(I)$ being a non-decreasing function in I with a non-increasing slope. ■

Corollary 1 *When a pure strategy equilibrium exists, it is unique.*

Proof. Suppose that t is the first time at which the equilibrium policies differ. If total quantities $\tilde{q}^t$, q^t, s.t. $\tilde{q}^t > q^t$, are both optimal solutions to the retailer's problem corresponding to the same equilibrium vector of prices $\tilde{w}^t = w^t$, then from Proposition 12 it follows that $DP(I^t + \tilde{q}^t - ED^t) > DP(I^t + q^t - ED^t)$. This contradicts the optimality of these policies. Suppose, on the other hand, that equilibrium price $\tilde{w}^t \neq w^t$. Then for the inactive suppliers, $q_i^t = \tilde{q}_i^t = 0$, implies that $\tilde{w}_i = w_i$. If $q_i^t \neq \tilde{q}_i^t = 0$, then it follows that $\tilde{w}_i > w_i$, and by Assumption 16, $\tilde{w}_i = s_i$. These two statements cannot hold simultaneously; hence, the sets of active and inactive suppliers are the same at time t. Since the prices set by the inactive suppliers, the demand and the initial inventory completely determine the equilibrium prices of active suppliers, these prices must be equal as well. Hence the equilibrium policy is unique. ■

Finally we observe that a solution to problems (3.19) and (3.20) indeed constitutes an MPNE.

Proposition 13 *An equilibrium policy is an MPNE.*

Proof. Suppose to the contrary that an equilibrium policy (i.e., a pair $(\mathbf{w}, \mathbf{q})$ that solves both (3.19) and (3.20)) is not a MPNE. Let us also denote by I the state variables corresponding to these policies. Then, if the subgame perfection property does not hold, for some period k, for some agent (a supplier or the retailer), the restriction of his/her strategy on the periods $[k+1 : T]$ is not optimal in the subgame starting at period k. But then this agent's overall strategy could be improved, if he/she uses an optimal strategy during the first $k-1$ periods and then uses a strategy that outperforms his/her equilibrium strategy on the last $T-k$ periods. The Markovian property must hold for every strategy, since at any period the strategies of the agents depend only on the current inventory level at the retailer. ■

Notice also that establishing of the existence of a pure strategy MPNE is similar to Theorem 8 (i.e., the continuous time case):

Theorem 9 *During each period t, the following condition is sufficient for a vector $(\mathbf{w}^*, \mathbf{q}^*)$ to constitute a pure strategy MPNE in the system:*

(1) $\frac{\partial \pi_i(\mathbf{w}^*)}{\partial w_i^t} \leq 0$ *and* $w_i^* \leq \min(s_{k+1}, p(D))$ *for every active supplier i,*

Together with the following condition this condition is also necessary for the existence of a pure strategy equilibrium:

(2) When the number k of active suppliers is greater than one, $q^t(\mathbf{w}^*) = K^k$.

3.4 Coordination

As we have already mentioned, it is easy to see that the centralized solution to our problem results in a higher overall profit in the system as well as a larger output. This observation motivates us to devise contracts that induce a behavior in the suppliers and the retailer that would increase the overall system's profit (which will be larger than in the decentralized setting). Furthermore, contracts can allow the agents to remain competitive. In this section we will first discuss the differences between the centralized and decentralized solutions and then we will suggest contracts that *coordinate* the system, i.e., the contracts that extract the centralized solution and associated profit from the system while maintaining competition. We will use superscript c in the notation related to the centralized solution and superscript d, to the decentralized solution. Notice that the total profit in the system at time t is completely determined by the vector of quantities ordered, $\mathbf{q}^t$, and could be expressed as $\pi^c(\mathbf{q}^t)$.

Example 3 Consider a two-supplier competition, as in Example 1, and suppose that condition $a - 2bK^2 \geq \max(\frac{a+s_i}{2} - bK_{-i}, s_i)$ holds. This condition is necessary for the existence of a pure strategy equilibrium. Then the centralized and the decentralized solutions coincide ($q_i = K_i$), i.e., the decentralized solution captures the system's optimal profit. On the other hand, if one of the suppliers, say supplier 1, is able to maintain a monopoly then in the centralized setting $q^c = q_1^c = \frac{a-s_1}{2b}$, whereas in the decentralized case $q^d = q_1^d = K_1, w_1 = s_2$, as long as $\frac{a-s_1}{2b} \leq K_1$ and $s_1 \leq s_2 \leq \frac{a+s_1}{2}$. Thus, the difference in profits would be $\pi(\mathbf{q^c}) - \pi(\mathbf{q^d}) = \frac{a-s_1}{2b} \cdot (\frac{a+s_1}{2b} - s_1) - K_1(s_2 - s_1) \geq 0$. This last inequality follows from the centralized problem formulation and is strict when $p(q)q$ is strictly concave ($b > 0$) and $a - s_1 \neq 2bK_1$.

Some properties in a single period case

Define $K_{-} = K - \varepsilon$ for some $\varepsilon > 0$. The lemma below describes some of the comparative properties of the centralized and decentralized solutions.

Lemma 7 *Suppose* $q^c \in [K^{k-1}, K^k]$, *then (a)* $q^d \leq q^c$, *(b)* $q^d \geq K^{k-2}$, *(c)* $q^d \in [K_{-}^{k-1}, K^k)$, *(d)* $w^d = s_{k+1}$, *(e)* $\pi^c(\mathbf{q}^c) \geq \pi_r(\mathbf{q}^d|\mathbf{w}^d) + \sum_{i=1}^{n} \pi_i(w_i^d|\mathbf{w}_{-i}^d)$.

Proof.

(a) $q^d \leq q^c \leq K^k$. These inequalities hold since $\pi_r(\mathbf{q}|\mathbf{w}, I)$ is increasing and $\mathbf{s} \leq \mathbf{w}$.

(b) $q^d \geq K^{k-2}$. Suppose the opposite holds. Then $s_{i+1} - R'(K^{i-1}) \leq 0$, $s_{i+2} - R'(K^i) > 0$, for some $i < k-2$, whereas $\forall i \leq k$, $s_i - R'(q^c) \leq 0$, with $q^c \in [K^{k-1}, K^k]$. Since $R'(q)$ is decreasing, $-R'(K^i) \leq -R'(q^c)$, $\forall i \leq k-1$. This implies that $s_{i+2} - R'(K^i) \leq s_{i+2} - R'(q^c)$. Notice that the right hand side expression is strictly negative $\forall i < k-2$. Hence $s_{i+2} - R'(K^i) < 0$, $\forall i < k-2$. This gives us a contradiction.

(c) Suppose $q^d < K^{k-1}$. From the previous item, it follows that $s_k - R'(q^d) = 0$ or $s_k < R'(K^{k-1}) < s_{k+1}$. The former is impossible, since $s_k - R'(q^c) < 0$ and R is a concave function. If the latter holds, then the optimal solution is $q^d = K^{k-1}$. On the other hand, notice that $w^d > s_k$ unless $q^d = K^{k-1}$, hence $q^c > q^d$.

(d) $w^d = s_{k+1}$ if $q^d \in [K^{k-1}, K^k)$ and $w^d = s_k$, if $q^d = K_{-}^{k-1}$, where K_{-}^{k} is a quantity slightly smaller than K^k.

(e) Consider $\delta = \frac{1}{2K}(\pi^c(q^c) - \pi^d(q^d))$. Since $q^c > q^d$, and R is strictly increasing, $\delta > 0$. ■

Coordinating mechanisms

We first point out that several types of contracts commonly considered in the literature will not work as coordinating mechanisms in our model. Among them are the following:

(1) Tolls set by the suppliers in such a way that they are proportional to the quantities ordered by the retailer. These contracts cannot work since a problem with tolls can be transformed into a problem without tolls, which would have the same double marginalization effects as those in the original model.

(2) Two-part tariffs set by the suppliers or the retailer in such a way that neither the retailer nor any of the suppliers is worse off than in the decentralized case.

For example, in case (2), consider a one-period game. Suppose that (L_i, w_i) are (fixed charge, per unit price) tariffs offered by the suppliers, and suppose that they constitute an equilibrium. Then $\sum_{i=1}^{n} L_i = \pi^c(q^c)$; otherwise, some supplier could have increased his/her profit through increasing his/her fixed charge. Moreover, prices and costs of active suppliers must be equal, i.e., $s_i = w_i = w_j = s_j$ if $q_i > 0$ and $q_j > 0$. Notice that otherwise, if $w_i > s_i$, then the total profit from the system would be lower than from a centralized system. And if $w_i > w_j$, then supplier j would benefit from increasing his/her price. In the single period case, it can be shown that the necessary conditions for two-part tariffs to be coordinating are (i) $\sum_{i=1}^{n} L_i = \pi^c(q^c)$, (ii) for all active suppliers in the centralized solution, $w_i = s_i = const$, (iii) for all the active suppliers in the centralized solution, $R(q^c) - R(q^c - q_i^c) - L_i - s_i q_i^c = const \geq 0$. This last condition comes into play when eliminating the possibility of an increased fixed charge by any active supplier while condition (i) holds.

Notice that these types of contracts are considered in the literature on coordination in the single-supplier/single-retailer case.

Tolls offered by the retailer in a single period case

Suppose the retailer announces tolls t_i that each supplier i needs to pay to the retailer. Then the suppliers announce prices w_i per unit of product. Based on these prices, the retailer determines the amount q_i to order from supplier i, and he/she pays $(w_i - t_i)q_i$ to that supplier, where t_i is the toll.

Suppose that the retailer sets tolls $t_i = w^d - s_k$, $\forall i$, where s_k is the highest cost of a supplier who is active in the centralized solution and w^d is the competitive price that the suppliers would have achieved if they knew that the retailer orders the centralized quantity q^c.

For example, when $q^d, q^c \in [K^{k-1}, K^k]$, the retailer charges extra $s_{k+1} - s_k$ per unit of the ordered quantity, and orders the total of q^c, from all the suppliers. When $q^d \in [K^{k-2}, K^{k-1})$, there is no need to impose tolls. The retailer will benefit from ordering q^c instead of q^d; incidentally in this case we can show that $q^d = K^{k-1} - \delta$, where δ is a very small positive number.

3.4.1 Coordination in the finite time horizon case

Assume that $\{D^t\}$ is a deterministic process.

Proposition 14 *There exist tolls* $\mathbf{t}$ *set by the retailer that coordinate the chain over the entire time horizon.*

Proof. We will prove this using backward induction. Suppose that during period t, in the centralized solution, k is the index of an active supplier with the highest marginal cost and $w^d(q^{c,t}) = w^d$ is equal to what the value of a decentralized equilibrium price would be if amount $q^{c,t}$ were ordered in the decentralized system. Our hypothesis is that the following tolls can be used to coordinate the chain

$$t_i^t = w^d - s_k - g^{c,t+1}(I^{c,t} + q^c - D^t) + g_r^{t+1}(I^{c,t} + q^c - D^t), \tag{3.22}$$

where $g(x,t) = DP'(x,t)$ denotes the derivative of the optimal profit-to-go starting at time t (index c corresponds to the centralized case while r to the retailer in the decentralized case).

First, the one-period case (time T) was shown to work in the previous subsection.

Suppose that the hypothesis holds at time $t+1$ and consider time t. Let I denote the inventory at time t in the centralized solution. The retailer optimizes the following objective:

$$R(q) - b(I + q - D)^- - h(I + q - D)^+ - (\mathbf{w} - \mathbf{t})'\mathbf{q} + DP_r(I + q - D, t + 1),$$

while the central manager optimizes

$$R(q) - b(I + q - D)^- - h(I + q - D)^+ - \mathbf{s}'\mathbf{q} + DP^c(I + q - D, t + 1).$$

When the optimal amount that the retailer orders in the problem with tolls equals that in the centralized problem, then tolls on active suppliers, set as $t_i = R'(q^c) - h + w_i + g_r(I + q^c - D, t + 1)$, coordinate the chain, when $I + q^c > D$ holds. Notice $t_i = R'(q^c) - b + w_i + g_r(I + q^c - D, t + 1)$, when $I + q^c < D$.

From the centralized solution, $R'(q^c) - h + s_k + g^c(I + q^c - D, t + 1) = 0$, hence, $t_i = -s_k + w_i - g^c(I + q^c - D, t + 1) + g_r(I + q^c - D, t + 1)$, where $w_i = w^d(q^c)$ since the suppliers will price competitively. Hence tolls set as in (3.22) are going to coordinate the subgame starting at period t with the inventory equal to the inventory in the centralized solution at time t. ■

3.5 Computations

In our computations we used examples of multi-period competition among several suppliers and a single retailer. We considered the VIP reformulation of the model for the retailer and the suppliers' problem. We used a serial version of the Decomposition algorithm for solving VIPs, (Gauss-Seidel method, see [46]), given that our feasible set can be partitioned into two sets: $(\mathbf{w}, \zeta) \in [\mathbf{s}, \infty) \times \mathbb{R}^{n+}$ and $\mathbf{q} \in [\mathbf{0}, \mathbf{K}]$. Below we present this algorithm. We use the notation $x = (\mathbf{w}, \zeta)$.

A decomposition algorithm to find the the decentralized solution

Step 0 Start with $x^0 = (\mathbf{w}^0, \zeta^0, \rho^0)$, some tolerance $\varepsilon > 0$. Set $k = 1$.

Step 1 Compute the solution $q^k = q$ to the $DVIP$:

$$F_q^I(q, x^{k-1})'(q' - q) \geq 0 \;\forall q' \in [0, \mathbf{K}].$$

Step 2 Compute the solution $x^k = x$ to the $DVIP$

$$F_x^I(q^k, x)'(x - x') \geq 0, \;\; \forall x' \in [\mathbf{s}, \infty) \times \mathbb{R}^{2+}.$$

Step 3 If $\|q^k - q^{k-1}\| \leq \varepsilon$, then stop. Else set $k \leftarrow k + 1$, and go to Step 1.

The $DVIP$ function is monotone when $p(D)$ is linear with decreasing slope. However, the feasible set is non-convex. Therefore, the convergence results from the literature for the Gauss-Seidel algorithm do not apply. Nevertheless, in the problem instances we considered, one of the major conditions for convergence, namely, monotonicity, is satisfied. Therefore, we believe that the algorithm is likely to converge.

We used MATLAB's linear programming solver to find a solution of the centralized version of the problem. We also considered a *myopic* solution of the problem, in which we solved problems ignoring intertemporal dependencies, i.e., just solving a one-period competition problem starting from time $t = 1$ and updating the next period's initial inventory.

In our computational examples, we investigated the following:

1. The decentralized vs. centralized solution and the difference in profits as a function of the number of competing suppliers and of variability in demand.

2. The decentralized solution vs. the myopic solution.

3. The profits as functions of the inventory.

4. The influence of variability in demand on prices, depending on the number of competitors.

5. The influence of capacity sizes: a large supplier vs. a small supplier (in terms of capacity).

6. The algorithm's convergence to a Nash equilibrium policy.

The difference between the decentralized solution and the centralized solution is small and is likely to decrease as the number of competitors increases. The variability in prices is likely to be amplified by the competition. This is due to the non-continuity of prices as functions of the quantities ordered. In Figure 3-4, we illustrate equilibrium prices corresponding to the demands drawn from a normal distribution with the same mean and different variations. More, precisely, the demand processes were simulated

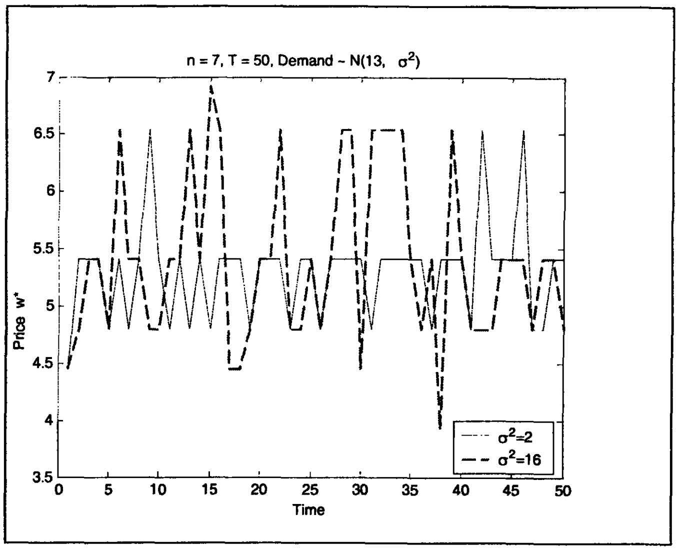

Figure 3-4: Changes in suppliers' prices $w^*(\mathrm{t})$ in time

using $D^t \sim \mathbf{N}(13, \sigma^2)$, where standard deviations are chosen to be $\sigma^2 = 2$ and $\sigma^2 = 16$.

We also compare centralized, decentralized and myopic solutions in a computational example, see Figure 3-5. The myopic solution solves the decentralized problem at each (single) period instead of solving it over the whole horizon. Thus during each period, starting with period 1, we know the values of the initial inventory and the consumers' demand. Based on these quantities, we compute equilibrium policies for that period. We then update the inventory variable and resolve the problem for the next period. We observe that the myopic solution is considerably worse than the decentralized solution in terms of the profit. The centralized solution is, as expected, the best, however the difference between the centralized and decentralized solution is not very large, it is about 6% in this particular example.

As Proposition 12 predicts, the retailer's profit is a piece-wise linear increasing function of the initial inventory, see Figure 3-1 when the initial inventory is smaller than the demand.

Another important parameter of the system is the ratio of the holding/backorder costs to the ordering costs. If holding/backorder costs are high then it makes no sense to carry inventory and , therefore, the problem becomes static. Finally in the

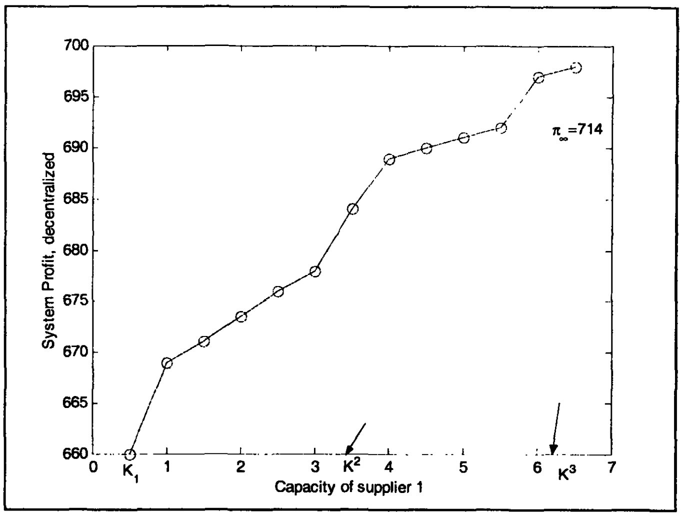

Figure 3-5: Total decentralized systems profit as function of a suppliers capacity

presence of a large supplier, fewer suppliers will participate and the equilibrium prices will be lower. In the next Figure 3-6 one can see that the profit is piecewise linear in terms of capacity of a supplier, the slopes decrease at points corresponding to the capacities in the system where marginal costs increase ($K^1 = K_1$, K^2, K^3 etc.).

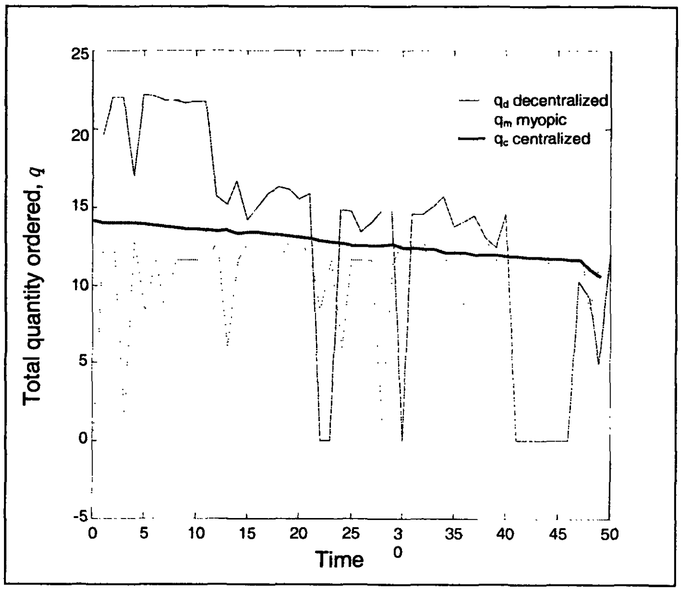

Figure 3-6: Total decentralized systems profit as function of a suppliers capacity

The numerical examples, we considered, illustrate some of the properties we study in this chapter. They also provide a further understanding of how strong the differences are among the various strategies of managing competition among several suppliers for the orders from a single buyer are in a dynamic setting. We can summarize our observations as follows:

- As expected, the centralized solution outperforms the decentralized one, and both substantially outperform the myopic solution in terms of the overall system's profit.
- Increasing the active supplier's capacity will lead to a piecewise linear increase in the decentralized system's profits.
- The difference between the decentralized and the centralized solution is likely to be non-increasing in the suppliers' capacities.
- The variability in the consumers' demand is amplified by the variability of the suppliers' capacities.
- The more the holding costs increase relatively to the suppliers' costs, the closer the decentralized solution is to the myopic solution.

3.6 A modified model with options

In this section we consider a decentralized setting with options on the suppliers capacities in the discrete time setting. There is a vast amount of literature on options, both in a financial context ([30]) as well as in other applications ([44]). The model in this section is the most related to the option models in [60] and [41], [42]. These models also involve a many-suppliers/one-retailer competitive setting. Nevertheless, there are some key differences with this work. [60] and [42] consider only a static setting. [41] concentrates on the retailer's problem (studying the costs involved rather than the properties of the equilibrium policies).

Suppose that in addition to n suppliers with fixed capacities, there is one supplier with unlimited capacity, whose price, p_x^t, fluctuates according to a cdf F_x for each period t. Consider a modified model of competition among suppliers. Suppose that at time 0 the suppliers offer options (o_i^t, v_i^t) on their capacity, where o_i^t is a reservation price per unit of capacity at time t and v_i^t is an exercise price per unit of capacity at time t, and $t = 1, \cdots, T$. Suppose that the retailer reserves Q_i^t units of capacity from supplier i at time 0, and, at time t, once he/she observes the outside supplier's price, he/she exercises $q_i^t \in [0, Q_i^t]$ units, then the retailer's total payment to suppliers at time t is: $\mathbf{o}^{t\prime}\mathbf{Q}^t + \mathbf{v}^{t\prime}\mathbf{q}^t + p_x^t x^t$, where $0 \le \mathbf{q}^t \le \mathbf{Q}^t$, $0 \le \mathbf{Q}^t \le \mathbf{K}$, $x^t \ge 0$. We also introduce an artificial supplier representing unsatisfied customers' demand with $0 \le q_{n+1}^t \le Q_{n+1}^t = D^t$, $o_{n+1}^t = 0$, and $v_{n+1}^t = 0$. We also assume that the demand process $\{D^t\}$ is exogenous and deterministic. Before studying this problem in a multi-period setting, we present properties of the model in a one-period case.

3.6.1 The one-period problem

The sequence of events in the one-period problem is illustrated in Figure 3-7.

Figure 3-7: Sequence of events in one-period options model

We first consider *the retailer's allocation problem.* The retailer solves this problem after reserving capacities $\mathbf{Q}$ and learning the price p_x of the outside supplier as well as the consumers' demand D. At this point he/she needs to decide how much product to purchase from each supplier (including the outside supplier), i.e. he solves the following problem:

$$\max_{\substack{0 \le \mathbf{q} \le \mathbf{Q},\ x \ge 0 \\ q + x \ge D - I}} \{r(q_{n+1}) - (h+b)q_{n+1} - (\mathbf{v} + h\mathbf{e})\mathbf{q} - (p_x + h)x - hI + hD - \mathbf{o}'\mathbf{Q}\}. \quad (3.23)$$

The solution to this problem is similar to that presented in (3.21) and can be expressed as follows:

$$q_i = \begin{cases} \min(Q_i, (D - I - \sum_{v_j < v_i} Q_j)^+) \frac{Q_i}{\sum_{v_j = v_i} Q_j}), & \text{if } v_i \le \min(p_x, p(D) + b); \\ 0, & \text{otherwise}; \end{cases}$$

$$x = (D - I - \sum_{v_j \le p_x} Q_j)^+, \text{ when } p(D) + b > p_x; \tag{3.24}$$

$$q_{n+1} = (D - I - \sum_{v_j \le p(D)+b} Q_j)^+, \text{ when } p(D) + b < p_x.$$

We note at this point that when $p(D) + b > p_x$, the consumers' demand will be fully satisfied and

$$D - I = \begin{cases} x + q, & \text{if } p(D) + b > p_x; \\ q, & \text{otherwise.} \end{cases} \tag{3.25}$$

Next we consider *the retailer's reservation problem.* At the time of reservation, the retailer knows the suppliers' offers $(\mathbf{o}, \mathbf{v}, \mathbf{K})$, but has not yet learn the price of the outside supplier. Thus, his/her reservation policy $\mathbf{Q}$ is an insurance against this outside price becoming too high. It is the solution of the following problem:

$$\max_{0 \le \mathbf{Q} \le \mathbf{K}} E_{p_x}\{r(q_{n+1}) - (h + b)q_{n+1} - (\mathbf{v} + h\mathbf{e})\mathbf{q} - (p_x + h)x\}, \tag{3.26}$$

where $\mathbf{q}, x$ are solutions of (3.23).

Using (3.25), this problem can be rewritten as

$$\max_{0 \le \mathbf{Q} \le \mathbf{K}} -E_{p_x}\{(\mathbf{v} - p_x\mathbf{e})'\mathbf{q} - r(q_{n+1}) + bq_{n+1}\}, \tag{3.27}$$

In order to solve this problem, we consider the following cases below.

(a) Suppose that relations $I \le D$ and $p_x < p(D) + b$ hold a.e.. Then $q_{n+1} = 0$ and $x + q = D - I$ a.e. Therefore, if all v_i's are distinct, then $q_i = 1_{v_i < p_x} Q_i 1_{\sum_{j:v_j < v_i} Q_j < D - I}$

and for an active supplier i, it holds that

$$E_x(v_i - p_x)q_i = Q_i E_x(v_i - p_x | v_i < p_x).$$

Let

$$u_i = E_x(v_i - p_x | v_i < p_x) = -E_x(p_x - v_i)^+,$$

then the entire objective becomes $(\mathbf{o} + \mathbf{u})\mathbf{Q} + hI + (E_x p_x - p(D))D$ and, hence,

$$\mathbf{Q}^* = \arg \min_{0 \leq \mathbf{Q} \leq \mathbf{K}} (\mathbf{o} + \mathbf{u})' \mathbf{Q},$$

i.e., $Q_i = K_i$, if $o_i + u_i \leq 0$ and $Q_i = 0$, otherwise.

(b) When $p_x > p(D) + b$ and $I < D$, then the retailer orders only from the capacitated suppliers. Moreover, $q = D - I$ and the retailer would seek to maximize $-(\mathbf{o} + \tilde{\mathbf{u}})'\mathbf{Q}$, where $\tilde{u}_i = -(p(D) + b - v_i)^+$. Hence, $Q_i = K_i$, if $o_i + \tilde{u}_i \leq 0$, and $Q_i = 0$, otherwise.

(c) Finally, when $I \geq D$, $\mathbf{Q} = 0$.

Because of the competition from other suppliers, supplier i will try to set his/her price v_i to the lowest level at which he/she remains competitive. Thus, the suppliers with smaller marginal costs have more bargaining power. In order to find an equilibrium solution, the retailer will place orders to the suppliers in the increasing order of their marginal costs and then will decide which capacities to reserve based on the cases (a)-(c) we presented above.

We finally consider *the suppliers' problem.* Supplier i solves the following problem:

$$\max_{o,v \geq 0} o_i Q_i(\mathbf{o}, \mathbf{v}) + E_x \left\{ q_i(\mathbf{o}, \mathbf{v}, p_x)(v - s_i) \right\}, \tag{3.28}$$

where $q_i(\cdot)$, $Q_i(\cdot)$ are determined from the retailer's problem.

To find solution to (3.28), observe that from the retailer's problem, supplier i can deduce that he/she needs to set parameters (o_i, v_i) so that $o_i + u_i = const$. For an active supplier i, it holds that $Q_i = K_i$ and $\frac{\partial o_i}{\partial v_i} = -\frac{\partial u_i(v_i)}{\partial v_i} = -F_x(v_i < p_x)$. Moreover,

$E_x\left(q_i(\mathbf{o},\mathbf{v},p_x)(v-s_i)\right) = K_i(v_i-s_i)F_x(p_x > v_i)$. Hence,

$$\frac{1}{K_i}\frac{\partial \pi_i(\mathbf{o},\mathbf{v})}{\partial v_i} = -F_x(p_x > v_i) + F_x(p_x > v_i) + (v_i - s_i)f_x(v_i).$$

Therefore, from the KKT conditions for (3.28), we obtain $v_i = s_i$ and $o_i = E_x(p_x - s_i)^+$. (Recall that all indices are increasing in the order of the suppliers' marginal costs.) Also observe that the solution, i.e, the capacities reserved by the retailer, as well as the reservation prices o_i and exercise prices v_i announced by the suppliers, can be viewed as a contract. This contract is equivalent to a two-part tariff with a fixed payment $L_i = K_iE_x(p_x - s_i)^+$ and a price $w_i = s_i$.

The case when index $\tilde{\mathbf{u}}$ is utilized will result in the same formulas for (o_i, v_i).

The following two results characterize the equilibrium in the single-period case.

Proposition 15 *Let $u_i^* = E_x(p_x - v_i^*)^+$ and suppose that a pure equilibrium strategy exists. The equilibrium strategy for the retailer is to reserve $Q_i^* = K_i$ when $o_i^* + u_i^* \leq 0$ and reserve nothing otherwise, and for the suppliers to announce $\mathbf{o}^* = E_x(p_x\mathbf{e} - \mathbf{s})^+$ and $\mathbf{v}^* = \mathbf{s}$.*

As a consequence, at an equilibrium, $o_i + u_i = 0$ for the active suppliers. Moreover, at the beginning of the period, the expected profits-per-period for the agents are as follows:

$$\pi_i(\mathbf{o}^*,\mathbf{v}^*) = K_i 1_{D-I\in(K^{i-1},K^i]}E_x(p_x - s_i)^+, \tag{3.29}$$

$$\pi_r(\mathbf{Q}^*) = Dp(D) - (D-I)E_x p_x, \tag{3.30}$$

$$\pi_o^d(I_0) = p(D) - (D-I)E_x p_x + \sum_{i=1}^{n} K_i 1_{D-I\in(K^{i-1},K^i]}E_x(p_x - s_i)^+, \tag{3.31}$$

where π_o^d is the total system's profit.

Notice that the equilibrium and the profits can be characterized using $\mathbf{u}$ instead of $(\mathbf{o},\mathbf{v})$. Let us use notation $\pi_i(\mathbf{u}) = \pi_i(\mathbf{o},\mathbf{v})$. Then the conditions for existence of pure equilibria can be stated as follows.

Theorem 10 *Let* $\mathbf{u}^*$ *(or* $\tilde{\mathbf{u}}^*$*) be as defined in this section. The following conditions are sufficient for a vector* $(\mathbf{o}^*, \mathbf{v}^*, \mathbf{Q}^*)$ *to be a pure strategy Nash equilibrium solution in the system:*

(1) $\frac{\partial \pi_i(\mathbf{u}^*)}{\partial u_i} \leq 0$, *for the suppliers with* $Q_i^* > 0$,

(2) $u_i^* = s_i$, *for the suppliers with* $Q_i^* = 0$.

The previous conditions are necessary and sufficient for the existence of a pure strategy equilibrium if in addition one of the following conditions hold:

(3a) $Q^*(\mathbf{o}^*, \mathbf{v}^*) = K^k$ *for some* k, *where* Q^* *is the total capacity reserved at the equilibrium,*

(3b) $p_x \leq p(D) + b$ *a.e.*

Moreover, the option contracts above coordinate the chain.

Proof. The proof of the necessity and sufficiency is similar to that for Theorem 8.

Notice that *the centralized* version of the model in this section has the objective as in (3.23) with $\mathbf{Q} = \mathbf{K}$. In the centralized case, the second stage problem's solution (after prices of the outside supplier are observed) is $q_i = K_i 1_{p_x > s_i} 1_{D-I \in (K^{i-1}, K^i]}$. It is obvious that the expected value of the solution is the same as in the case when option contracts are used in the system and the expected total profit from the system is:

$$\begin{aligned} \pi_{co} =& E_x \left(p(D)(D - q_{n+1}) - h(I + q + x - D) - bq_{n+1} - \mathbf{s}'\mathbf{q} - p_x x\right) && (3.32) \\ =& Dp(D) - 2hD - hI - E_x p_x D + E_x(p_x - s_i) K_i 1_{D-I \in (K^{i-1}, K^i]} && (3.33) \\ =& \pi_{do}. && (3.34) \end{aligned}$$

When $I < D$, the first equality holds and the second equality follows from the decentralized version of the problem. When $I \geq D$ in both the centralized and the decentralized problems no additional orders will be placed. ■

3.6.2 Multi-period problem

In this subsection we will consider properties of the agents' policies in a multi-period case. A pure strategy MPNE exists when conditions similar to those in Theorem 10 apply for each period.

Theorem 11 *Let* $\tilde{u}_i^t = -E_x 1_{p_x<p(D^t)+b}(p_x - v_i^t)^+ - (1-F_x(b+p(D^t))(b+p(D^t)-v_i^t)^+$ *and suppose that a pure strategy MPNE exists. The equilibrium strategy for the retailer is to reserve* $Q_i^t = K_i$ *when* $o_i^t + u_i^t \leq 0$ *and reserve nothing otherwise, and for the suppliers to announce* $\mathbf{o}^* = E_x(p_x\mathbf{e} - \mathbf{s})^+$ *and* $\mathbf{v}^t = \mathbf{s}$*.*
This equilibrium strategy coordinates the supply chain.

Proof. We will prove by backward induction. As was shown in Lemma 15, this theorem holds for the single period case. Therefore, it holds for period T.

Consider period k, $1 < k < T$. Let us denote by $DP_{ro}^k(I)$ the value of the retailer's profit restricted to the last $T-k+1$ periods. This function is piecewise linear (see Proposition 12). According to the Proposition 12, the retailer's profit-to-go function $DP_{ro}^k(I)$ can be expressed as $DP_{ro}^k(I) = \gamma_0(I) + \gamma_I(I)I$, where coefficient $\gamma(I)$ is non-negative and non-increasing.

During period k, the retailer's objective at the allocation stage is

$$DP_o^k(I) = \max_{\substack{0\leq \mathbf{q}\leq \mathbf{Q},\\ x\geq 0,\\ q+x\geq D-I}} \Big((p(D)-b)q_{n+1} - (\mathbf{v}+h\mathbf{e}-\gamma_I(I^+)\mathbf{e})\mathbf{q} - (p_x+h-\gamma_I(I^+))x$$
$$+p(D)D - (h-\gamma_I(I^+))(D-I) + \gamma_0 - \mathbf{o}'\mathbf{Q}\Big), \quad (3.35)$$

where $I^+ = q+x+I-D$. This is a linear program and, if all the variables' coefficients are nonnegative, then the solution is analogous to that in (3.24). Moreover, notice that either $x=0$ or $q_{n+1}=0$, depending on the sign of expression $p_x - b - p(D)$, and only the relative values of v_i, p_x and $p(D)+b$.

Now suppose that $I^t < D^t$, then the reservation problem during period k could be written similar to the one in (3.26):

$$\max_{0 \leq \mathbf{Q} \leq \mathbf{K}} \gamma_0(I) + (p(D) + h - \gamma_I(I^+))D - (h - \gamma_I(I^+))I - \mathbf{o}'\mathbf{Q}$$
$$- E_{p_x \leq b+p(D)}\{(\mathbf{v} - p_x \mathbf{e})'\mathbf{q} + (D - I)(h + p_x - \gamma_I(I^+)\}$$
$$- E_{p_x \geq b+p(D)}\{(\mathbf{v} - (b + p(D))\mathbf{e})'\mathbf{q} - (D - I)(h + b + p(D) - \gamma_I(I^+))\}, \quad (3.36)$$

where $(\mathbf{q}, x, y)$ are as defined in (3.24) (for the kth period's values).

Hence, the retailer will distribute the orders among n suppliers using index $\tilde{u}_i = o_i + E1_{p_x < b+p(D)}(p_x - v_i)^+ + (1 - F_x(b + p(D))(b + p(D) - v_i)^+$. Consequently,

$$\frac{\partial o_i}{\partial v_i} = F_x(b + p(D)) - F_x(v_i)$$

and as in the previous subsection it follows that at time k: $v_i = s_i$. Moreover, since at equilibrium $o_i + \tilde{u}_i = 0$, it follows that $o_i = -E1_{p_x < b+p(D)}(p_x - s_i)^+ - (1 - F_x(b + p(D))(b + p(D) - s_i)^+$.

So far we showed that if the induction hypothesis holds for the last $T - k$ periods and the level of initial inventory of time k, I^k, induces a pure strategy MPNE in the game truncated at time k, the suppliers and retailer's equilibrium strategies are as described in the theorem's statement. Now it is easy to see that these strategies would also coordinate the chain. Indeed,

$$\begin{aligned} DP^{c,t}(I^t) &= E_x\left(p(D)(D - q_{n+1}) - h(I^{t^+}) - bq_{n+1} - \mathbf{s}'\mathbf{q} - p_x x + DP^{c,t+1}(I^{t^+})\right) \\ &= \pi_c(I, D) + DP^{c,t+1}(I^{t^+}) \\ &= \pi_{do}(I, D) + DP_o^{t+1}(I^{t^+}) \text{ (using induction hypothesis)} \\ &= DP_o^t(I^t) \end{aligned}$$

This concludes the proof. ∎

Remarks.

1. Notice that with a different sequence of events, the full coordination of the chain with options would be impossible. Namely, suppose that retailer can reserve capacities not only at the time 0, but also at any other period. Then, since there is a tow-state problem during each period , the closed-loop options solution will always be better than a centralized open-loop solution.

2. Another possibility is to consider a modified model in which we allow the suppliers to continue to compete at stage 2 of each period. Then at the allocation stage, the retailer orders additional amounts from supplier i at price w_i if $w_i < p_x$, and $D - I - (\sum_{j:v_i \leq w_i} Q_j - \sum j : w_j < w_i) > 0$. In this model options might coordinate the chain.

3.7 Extensions: expandable capacities and nonlinear costs

In this section we discuss some extensions of the model we discussed so far.

3.7.1 Expandable capacities

Consider a model in which during period t each supplier starts with an initial capacity $K_{i,0}^t$ and has an option of expanding his/her capacity up to $K_{i,1}^t$ at a cost ζ_i^t. The total cost of producing q_i^t units of product is $s_i^t(q) = s_i q + \zeta_i^t(q - K_{i,0}^t)^+$, where $0 \leq q \leq K_{i,1}^t$. Suppose also that each supplier i announces to the retailer a price schedule $w_i^t(q) = w_i^t q + v_i^t(q - K_{i,0}^t)^+$, where $0 \leq q \leq K_{i,1}^t$. In this setup, we can show that the equilibrium solution is equivalent to that of a problem with fixed capacities, that is defined as follows: the model consists of a set of $2n$ suppliers, with costs $\tilde{s}_i^t$, s.t. $\tilde{s}_i^t = s_i$, $\tilde{s}_{i+n}^t = s_i + \zeta_i^t$ and capacities $\tilde{K}_i^t = K_{i,0}, \tilde{K}_{i+n} = K_{i,1} - K_{i,0}$, for $1 \leq i \leq n$. Let us denote the suppliers' prices and quantities in this new model as $\tilde{w}_i$s and $\tilde{q}_i$s.

In order to see that the problems are equivalent, we first observe that the retailer's problem can be formulated as

$$\max_{\mathbf{q}^t, \mathbf{y}^t \geq 0} \left\{ \sum_{t=1}^{T} \left(r(D^t - I^t + q^t + y^t) - \mathbf{w}^{t\prime}\mathbf{q}^t - \mathbf{v}^{t\prime}\mathbf{y}^t - hI^{t+} - bI^{t-} \right) \right.$$
$$\left. \mid \mathbf{q}^t \leq \mathbf{K}_0^t,\ \mathbf{y}^t \leq \mathbf{K}_1^t - \mathbf{K}_0^t, I^{t+1} = (I^t + q^t + y^t - D^t)^+, I^1 = I_0 \right\}. \quad (3.37)$$

Therefore, the retailer's problem is equivalent to a problem with $2n$ suppliers whose prices and capacities are $\mathbf{w}, \mathbf{K}_0$ and $\mathbf{v}, \mathbf{K}_1 - \mathbf{K}_0$. Supplier i's problem can also be formulated as follows:

$$\max_{w_i^t, v_i^t \geq 0} \sum_{t=1}^{T} \left((w_i^t - s_i) q_i^t(\mathbf{w}^t, \mathbf{v}^t) + (v_i^t - s_i - \zeta_i^t) y_i^t(\mathbf{w}^t, \mathbf{v}^t) \right). \tag{3.38}$$

Here q_i^t is the amount that is ordered from the first $K_{i,0}^t$ units of capacity and y_i^t is the amount that is ordered from the last $K_{i,1}^t - K_{i,0}^t$ units, left in supplier i's capacity. During each period this problem is separable in terms of variables v_i and w_i. Therefore, each supplier with a variable capacity can be modelled as two suppliers with fixed capacities $K_{i,0}$ and $K_{i,1} - K_{i,0}$ and costs s_i and $s_i + \zeta_i^t$.

So far we have assumed that extra capacity is available at each period. Nevertheless, using it does not affect the initial level of capacity that is available in the next period. This is due to the fact that when the next period capacity is set as

$$K_{i,0}^{t+1} = K_{i,0}^t + y_i^t, \tag{3.39}$$

supplier i's problem remains separable. On the other hand, with this capacity equations, the retailer's problem has an additional constraint on capacities, namely, $q_i^t \leq K^{t-1} + y^{t-1}$. Since this type of constraint is linear, the properties of the solution to the retailer's problem remain the same. Therefore, the model with expandable capacities (i.e., where (3.39) holds) can still be reformulated as a $2n$-supplier/1-retailer decentralized supply chain competition problem.

Finally, suppose that there is some additional constant charge θ_i^t that supplier i incurs whenever he makes the decision to expand capacity, i.e., supplier i's cost is $s_i^t(q) = s_i q + \zeta_i^t (q - K_{i,0}^t)^+ + 1_{q > K_{i,0}^t} \theta_i^t$, where $0 \leq q \leq K_{i,1}^t$. In this case, supplier i can again be split into two suppliers: one that seeks to maximize $(w - s_i) q_i(\mathbf{w}, \mathbf{v})$ and another supplier, maximizing $(v - \zeta_i^t - s_i) y_i(\mathbf{w}, \mathbf{v}) - \theta_i^t$. Solving these two maximization problems will allow supplier i to set his/her equilibrium price v high enough so the capacity expansion is profitable.

3.7.2 Piecewise linear and nonlinear costs

The following proposition summarizes and generalizes the discussion in this section.

Theorem 12 *The model in which suppliers have bounded piecewise linear increasing costs is equivalent to a model with fixed capacities.*

The case of piecewise linear increasing costs can be generalized to the case where costs are piecewise convex increasing functions such that their slopes are increasing in the intervals where the functions are continuous. When these types of costs are used and the retailer's revenue is a convex function. Then function $q(\mathbf{w}(\cdot))$ is well defined (through the retailer's problem) and, therefore, the suppliers' problems are well defined and an equilibrium exists (nevertheless, we need additional conditions for the existence of a pure strategy equilibrium).

Finally, we note that the convex costs assumption might not hold. For example, given that there are economies of scale, the costs should generally decrease as the quantities rise. However, when a supplier needs to install new capacities/technologies, the costs might be convex at least for a short period of time, and our analysis would provide a framework for studying such situations.

3.8 Conclusions

In this chapter we introduced and studied models of competition between several suppliers and a retailer in a supply chain. The suppliers have capacities and face constant costs. We first study the problem of dynamic competition using a continuous time framework. We then established properties of the equilibria and the profit functions in the discrete time case. To achieve a more efficient yet competitive solution, we introduced contracts to the system. We also analyzed an alternative model with options that improves the system's coordination and yet incorporates competition. In addition, we studied several extensions to include more general cost structures for the suppliers. Finally we studied the behavior of the system using a matlab simulation. We observed that the properties of the system that we established theoretically, indeed hold in the simulation. Moreover, computational examples we considered supported our thesis about the importance of the dynamic aspect of the chain, as the solutions that take into account the system's dynamics significantly outperform the static (myopic) solution. Moreover, the numerical examples reaffirmed that a further improvement can be achieved when the difference between the centralized and decentralized system's solutions is captured using contracts.

This research also gives rise to several open questions.

- Can we quantify the loss of efficiency in the chain due to competition?
- How can we extend our model to incorporate incomplete information?
- How can we consider a more general supply chain with several retailers but also multiple echelons?

Bibliography

[1] D. Barnes-Schuster, Y. Bassok, and R. Anupindi, 2000, "Coordination and flexibility in supply chain contracts with options," preprint.

[2] D. Beil and L. Wein, 2001, "An inverse-optimization-based auction mechanism to support a multi-attribute RFQ process," submitted to Management Science.

[3] F. Bernstein and A. Ferdegruen, 2001, "A general equilibrium model for decentralized supply chains with price and service competition," preprint.

[4] D. Bertsekas, 1995, *Dynamic programming and optimal control,* Athena Scientific.

[5] D. Bertsimas and I.Ch. Paschalidis, 2001, "Probabilistic service level guarantees in make-to-stock manufacturing system," Operations Research 49: 119-133.

[6] G.P. Cachon, 1998, "Competitive supply chain inventory management," in S. Tayur, R. Ganeshan and M. Magazine (eds.), *Quantitative models for supply chain management,* Kluwer Academic Publishers.

[7] G.P. Cachon and M. Larivier, 2000, "Supply chain coordination with revenue sharing: strengths and limitations," University of Pennsylvania working paper.

[8] G.P. Cachon and M. Fisher, 2000, "Supply chain inventory management and the value of shared information," Management Science, 46: 1032-48.

[9] G.P. Cachon, 2001, "Supply chain coordination with contracts," preprint.

[10] G.P. Cachon and P.T. Harker, 2002, "Competition and outsourcing with scale economier," Management Science 48: 1314-1333.

[11] G.P. Cachon and S. Netessine, 2004, "Game theory and supply chain analysis," in D. Simchi-Levi, S.D. Wu and Zh.-J. Shen (eds.), *Handbook of quantitative supply chain analysis: modeling in the e-business era,* Kluwer Academic Publishers.

[12] R. Caldentey and L. Wein, 1999, "Analysis of a decentralized production-inventory system," preprint.

[13] L.M.A. Chan, D. Simchi-Levi, and J. Swann, 2001, "Dynamic pricing strategies for manufacturing with stochastic demand and discretionary sales," preprint.

[14] R.R. Chen, G. Jamakiraman, R. Roundy and R.Q. Zhang, 2002, "Efficient auction mechanisms for supply chain procurement," submitted to Operations Research.

[15] F.Y. Edgeworth, 1925, *Papers related to political economy I,* McMillan.

[16] P. Dasgupta and E. Maskin, 1986, "The existence of equilibrium in discontinuous economic games I: theory," Review Economic Studies 53: 1-26.

[17] S. Dempe, 2002, *Foundations of bilevel programming,* Kluwer Academic Publishers.

[18] M. Denault and J.-L. Goffin, 1998, "Variational inequalities with quadratic cuts," preprint.

[19] I. I. Dikin, 1967, "Iterative solution of linear and quadratic programming," Soviet Mathematics Doklady 8: 674-675.

[20] E. Dockner, S. Jørgensen, N.V. Long and G. Sorger, 2000, *Differential games in economics and social science,* Cambridge University Press.

[21] K. Ertogral and S.D. Wu, 2001, "A bargaining game for supply chain contracting," preprint.

[22] M. Frank and P. Wolfe, 1956, "An algorithm for quadratic programming," Naval Research Logistics Quarterly 3: 95-110.

[23] D. Fudenberg and J. Tirole, 1986, *Dynamic models of oligopoly* in Fundamentals of pure and applied economics 3, Harwood Academic Publishers.

[24] D. Fudenberg and J. Tirole, 1991, *Game theory,* MIT Press.

[25] J. Gallien and L. Wein, 2000, "Design and analysis of a smart market for industrial procurement," submitted to Management Science.

[26] J. L. Goffin, P. Marcotte, and D. Zhu, 1997, "An analytic center cutting plane method for pseudomonotone variational inequalities," Operations Research Letters 20: 1-6.

[27] J.L. Goffin and J.-P. Vial, 1999, "Convex nondifferentiable optimization: a survey focused on the analytic center cutting plane method," Logilab Technical Report.

[28] C. C. Gonzaga and L. A. Carlos, 1990 (Rev. 2002), "A primal affine-scaling algorithm for linearly constrained convex programs," preprint.

[29] R. F. Hartl, S. P. Sethi, adn R.G. Vickson, 1995, "A survey of the maximum princeples for optimal control problems with state constraintm," SIAM Review 37: 181-218.

[30] J. C. Hull, 1989 *Options, futures, and other derivative securities,* Prentice Hall .

[31] M. Jin and S.D. Wu, 2001, "Supply chain coordination in electronic markets: auction and contracting mechanisms," preprint.

[32] M. Jin and S.D. Wu, 2001, "Modeling capacity reservation in high-tech manucturing," preprint.

[33] S. Jørgensen and P.M. Kort, 2002, "Optimal pricing and inventory policies: Centralized and decentralized decision making," European Journal of Operational Research 138: 578-600.

[34] D. Kreps and J. Scheinkman, 1983, "Quantity precommitment and Bertrand competition yield Cournot outcomes," Bell Journal of Economics 14: 326-337.

[35] M. Larivier, 1999, "Supply chain contracting and coordination with stochastic demand," in S. Tayur, R. Ganeshan, and M. Magazine (eds.), *Quantitative Models of Supply Chain Management*, Kluwer Academic Publishers.

[36] R. Levitan and M. Shubik, 1972, "Price duopoly and capacity constraints," International Economics Review 13: 111-122.

[37] H.J. Lüthi, 1985, "On the solution of variational inequalities by the ellipsoid method," Mathematics of Operations Research 10: 515-522.

[38] H.J. Lüthi and B. Büeler, 2000, "The analytic center quadratic cut method (ACQCM) for strongly monotone variational inequality problems," SIAM Journal of Optimization 2: 415-426.

[39] T. Magnanti and G. Perakis, 1995, "A unifying geometric solution framework and complexity analysis for variational inequalities," Mathematical Programming 71: 327-351.

[40] T. Magnanti and G. Perakis, 1998, "From Frank-Wolfe to steepest descent; a descent framework for solving asymmetric variational inequalities," MIT Working Paper.

[41] V. Martínez-de-Albéniz and D. Simchi-Levi, 2002, "A portfolio approach to procurement contracts," preprint.

[42] V. Martínez-de-Albéniz and D. Simchi-Levi, 2003, "Competition in the supply chain option market," preprint.

[43] F. S. Mokhtarian and J.-L. Goffin, 2000, "An analytic center quadratic cut method for convex quadratic feasibility program," preprint.

[44] W.T. Moore, 2001 *Real options and option-embedded securities,* John Wiley.

[45] A. Nagurney, J. Dong and D. Zhang, 2002, "A supply chain network equilibrium model," Transportation Research Part E: 281:303.

[46] A. Nagurney, 1993, *Network economics: A variational inequality approach*, Kluwer Academic Publishers.

[47] Y. Nesterov and A. Nemirovskii, 1994, "Interior Point Polynomial Algorithms in Convex Programming," SIAM Studies in Applied Mathematics 13.

[48] Y. Nesterov and J.-P. Vial, 1999, "Homogeneous analytic cneter cutting plane methods for convex problems and variational inequalities," SIAM Journal of Optimization 3: 707-728.

[49] J.-S. Pang and M. Fukushima, 2003, "Quasi-variational inequalities, generalized Nash equilibriua and multi-leader-follower games," preprint.

[50] J.-S. Pang, 2002, "Computing generalized Nash equilibria," preprint.

[51] G. Perakis and M. Zaretsky, 2002, "On the effiecient solution of variational inequalities; complexity and computational efficiency," preprint.

[52] S.P. Sethi and G.L. Thompson, 2000, *Optimal control theory - Applications to management science and economics,* Kluwer Academic Publishers.

[53] S. Spinler, 2003, *Capacity reservation for capital-intensive technologies - An options approach* , Springer-Verlag.

[54] J. Sun, 1996, "A convergence analysis for a convex version of Dikin algorithm," Annals of Operations Research 62: 357-374.

[55] J. Tirole, 1988, *The theory of industrial organization,* MIT Press.

[56] B. Tomlin, 2000, "Supply chain design: capacity, flexibility and whole sale price strategies," MIT Thesis.

[57] P. Tseng, 2001, "Convergence properties of Dikin's affine scaling algorithm for nonconvex quadratic minimization," preprint.

[58] R.J. Vanderbei and J.C. Lagarias, 1990, "I.I. Dikin's convergence result for the affine-scaling algorithm," in Contemprorary Mathematics 114: 109-119.

[59] D.J. Wu, P.R. Kleindorfer, Y. Sun, and J.E. Zhang, 2002, "Optimal bidding and contracting strategies for capital-intensive goods," European Journal of Operational Research 137: 657-676.

[60] D.J. Wu, P. R. Kleindorfer, and Y. Zhang, 2002, "B2B exchanges and flexible supply chain contracting," submitted to Operations Research.

[61] Y. Ye, 1989, "An extension of Karmakar's algorithm and the trust region method for quadratic programming," in N. Megiddo (ed.), *Progress in Mathematical Programming*, Springer.

[62] Y. Ye, 1992, "On affine scaling algorithms for nonconvex quadratic programming," Mathematical Programming 56: 285-300.

[63] P.H. Zipkin, 2000, *Foundations of inventory management*, McGraw-Hill.

Appendix A

A result for Theorem 7

Proposition 16 *Suppose Assumptions 3 and 12 hold. Then* $F(x^k) - \bar{F} \geq c\|d^k\|^2$ *for some* $c > 0$.

Proof. Paper [54] has shown that when the problem function f has the property of Jacobian similarity, then for some positive integer $n_1 \leq n$, there is an orthogonal matrix M, such that for any $x \in K$, it holds that $M'HM = \begin{pmatrix} Q_x & 0 \\ 0 & 0 \end{pmatrix}$, where $H = \nabla f(x)$ and matrix Q_x is an $n_1 \times n_1$ bounded, positive definite matrix. Therefore, if we let $H^k = \nabla f(x^k)$, then

$$M'H^kM = \begin{pmatrix} Q^k & 0 \\ 0 & 0 \end{pmatrix}.$$

We can express $d^k = M \begin{pmatrix} y^k \\ z^k \end{pmatrix}$ or $d^k = s^k + w^k$, where $s^k = M \begin{pmatrix} y^k \\ 0 \end{pmatrix}$ and $w^k = M \begin{pmatrix} 0 \\ z^k \end{pmatrix}$. From the proof of Proposition 5 it follows that

$$F(x^k) - F(x^{k+1}) \geq \left(\alpha^k\right)^2 \frac{\rho}{2} d^k \nabla f(\hat{z}) d^k,$$

therefore

$$F(x^k) - F(x^{k+1}) \geq \frac{(\alpha^k)^2}{2} y^k Q^0 y^k.$$

When $\alpha^k \geq \alpha$,

$$\|s^k\| = \|y^k\| \leq \frac{\mu_1}{\alpha} \left(F(x^k) - F(x^{k+1})\right)^{1/2}.$$

Next we show that $\|w^k\| \leq \frac{\mu_2}{\alpha} \left(F(x^k) - F(x^{k+1})\right)^{1/2}$, for some constant μ_2. Suppose this is not true. Then, since $\{w^k\}$ is a bounded sequence, there exists a subsequence S and a subset J of $\{1, ..., n\}$ so that

$$\left\{\frac{|F(x^{k+1}) - F(x^k)|^{1/2}}{\|w^k\|}\right\}_S \to 0, \quad \lim_{k\to\infty, k\in S} \frac{w_j^k}{\|w^k\|} > 0, \ \forall j \in J, \quad \lim_{k\to\infty, k\in S} \frac{w_j^k}{\|w^k\|} = 0, \ \forall j \in J^c.$$

Then from the properties we established above for $\|s^k\|$, it follows that

$$\left\{\frac{\|s^k\|}{\|w_j^k\|}\right\}_S \to 0, \ \forall j \in J.$$

Consider the following system of linear equations

$$\begin{cases} A(s^k + w) = 0 \\ f(x^0)'w = f(x^0)'w^k \\ w_j = w_j^k \ \forall j \in J^c \\ (M'w)_j = 0 \ \forall j = 1, ..., n_1. \end{cases}$$

Next we will show that there exists a solution of the above system bounded by $|F(x^k) - F(x^{k+1})|^{1/2}$. To achieve this it suffices to show that in the system above the norm of the left-hand-side is bounded by $O(|F(x^k) - F(x^{k+1})|^{1/2})$, and the matrix on the right-hand-side is bounded. Paper [54] showed that for any system $Ax \leq b$ that has a feasible solution, there exists some solution whose norm is bounded by $\lambda\|b\|$, where λ is a constant dependent only on matrix A. Notice that

$$\begin{aligned} F(x^k) - F(x^{k+1}) &= \alpha^k f(x^k)'d^k + {\alpha^k}^2/2\, {d^k}'\nabla f(z^k) d^k \\ &= \alpha^k f(x^0)'(w^k + s^k) + \alpha^k (x^k - x^0)'\nabla f(z_0^k) d^k + {\alpha^k}^2/2\, {d^k}'\nabla f(z^k) d^k. \end{aligned}$$

From this relation it follows that

$$\begin{aligned}-f(x^0)'w^k &= f(x^0)'s^k + \frac{1}{\alpha^k}\left(F(x^k) - F(x^{k+1})\right) \\ &+(x^k - x^0)'M\begin{pmatrix} Z_0^k & 0 \\ 0 & 0\end{pmatrix}\begin{pmatrix} y^k \\ z^k\end{pmatrix} + \alpha^k y^{k\prime} Z^k y^k,\end{aligned}$$

where $\|Z_0^k\|$, $\|Z^k\|$, $\|x^k - x^0\|$ are bounded on K, and $\alpha \le \alpha_k \le 1$. Therefore, since there exists a solution of this system, namely w^k, there exists a solution $\bar{w}^k$ that is of order $O\left(|F(x^k) - F(x^{k+1})|\right)$. Therefore $\{\|\bar{w}^k\|/\|w^k\|\}_S \to 0$. Let $t^k = w^k - \bar{w}^k$.

We now can estimate $\lim_{k\to\infty}\left(\frac{d_i^k - t_i^k}{d_i^k}\right)^2$:

$$\text{for } i \in J: \quad \lim_{k\to\infty}\left(\frac{d_i^k - t_i^k}{d_i^k}\right)^2 = \lim_{k\to\infty}\sum_{i\in J}\left(\frac{\frac{s_i^k}{\|w^k\|} + \frac{\bar{w}_i^k}{\|w^k\|}}{\frac{s_i^k}{\|w^k\|} + \frac{w_i^k}{\|w^k\|}}\right)^2 = 0, \qquad \text{(A.1)}$$

$$\text{for } i \in J^c: \quad \frac{d_i^k - t_i^k}{d_i^k} = 1.$$

The following system holds as well:

$$\begin{cases} (M't^k)_j = 0, \; j = 1, ..., n_1 \\ At^k = 0 \\ f(x^k)'t^k = f(x^0)'t^k + (x^k - x^0)'M\begin{pmatrix} Z^{t^k} & 0 \\ 0 & 0\end{pmatrix} M't^k = 0 \\ H^k t^k = M\begin{pmatrix} Q^k & 0 \\ 0 & 0\end{pmatrix} M't^k = 0.\end{cases}$$

Observe that for large $k \in S$, $d^k - t^k$ is an optimal solution of AS direction finding problem lying in the interior of $D(x^k)$. Notice that $f(x^k)'(d^k - t^k) = f(x^k)'d^k$, $A(d^k - t^k) = 0$ and from (A.1)

$$\|X^{k^{-1}}(d^k - t^k)\| = \sum_1^n \left(\frac{d_i^k - t_i^k}{d_i^k}\right)^2\left(\frac{d_i^k}{x_i^k}\right)^2 < \sum_1^n\left(\frac{d_i^k}{x_i^k}\right)^2.$$

This is a contradiction, since if there were an interior solution of direction finding problem, then the current iterate would be an optimal solution. ■

Appendix B

An extension of the generalized geometric framework

B.1 Definition of the framework

In this appendix we present an extension of the generalized geometric framework (GGF) originally proposed in [39]. Before presenting an extended version of GGF, we will state some definitions.

Definition 11 ([39]) "Nice" sets $\{P^k\}_0^\infty$ are subsets of $\mathbb{R}^n$ satisfying the following conditions:

1. the sets P^k are compact convex subsets of $\mathbb{R}^n$ with nonempty interior;
2. the sets P^k belong to the same class of sets (ellipsoids, polytopes, simplices, convex sets);
3. these sets satisfy the condition $P^{k+1} \supseteq P^k \cap H_k^c$ for a half space H_k^c defined by a hyperplane H_k;
4. we choose the set P^0 so that it contains the set K and has a volume no more than $2^{n(2L+1)}$.

Definition 12 ([39]) A sequence of point $\{x^k\}_0^\infty \subseteq \mathbb{R}^n$ is a sequence of *centers* of the "nice" sets $\{P^k\}_0^\infty$ if

1. $x^k \in \text{interior} P^k$;

2. $P^k \supseteq P^{k-1} \cap H_k^c = P^{k-1} \cap \{x \in \mathbb{R}^n \mid c^{k-1'}(x - x^{k-1} \geq 0\}$ for a given vector c^{k-1};

3. $\text{vol}(P^k) \leq b(n)\text{vol}(P^{k-1})$ for a function $b(n)$, $0 < b(n) < 1$, of the number n of the problem variables.

Definition 13 For some $g : \mathbb{R}^n \to \mathbb{R}$, *hypersurface* $\bar{S}$ is defined as follows: $\bar{S} = \{x \in \mathbb{R}^n | g(x) = 0\}$. We will say that a set § *lies above* the surface $\bar{S}$ if $S = \{x \in \mathbb{R}^n | g(x) \geq 0\}$ and that a hypersurface $\bar{S}(x)$ *supports* a set K if $K \subseteq S$.

We will present a framework that uses hypersurfaces as cuts, whereas linear cuts were used in [39].

An Extension of the General Geometric Framework (EGGF)

1. Start with an interior point x^0 - center of a "nice" set $P^0 \supseteq K^0 = K$.
2. Feasibility cuts. At iteration k, given P^k, K^k, and x^k (center of P^k):
   (i) Compute a surface $\bar{S}\left(x^k\right)$ that supports K. We will denote by $S\left(x^k\right)$ the set lying above this surface.
   (ii) Update P^{k+1}, so that $P^{k+1} \supseteq S(x^k) \cap P^k$, $K^{k+1} = K^k$.
3. Optimality cuts. At iteration k, given P^k, K^k, and x^k (center of P^k):
   (i) Compute a surface $\bar{S}\left(x^k\right)$ that cuts K^k through x^k. Moreover, the surface is such that the set $S\left(x^k\right)$ lying above this surface, contains all the VIP (or $WVIP$) solutions.
   (ii) Update $K^{k+1} = K^k \cap S(x^k)$, $P^{k+1} \supseteq S(x^k) \cap P^k$.
4. $k \leftarrow k + 1$. Repeat steps 2 and 3 until a desirable precision is reached.

In Magnanti and Perakis [39], the surface $\bar{S}(x^k)$ is a hyperpane $H(x^k)$ determined through a linear cut with slope $f(x^k)$. In Section 2.2 of this thesis, we consider alternate choices for the slopes of linear cuts.

The extension of the GGF allows us to incorporate nonlinear cuts. For example, an obvious but perhaps not very practical choice of a cut, could be to introduce at each step k, a nonlinear cut $S(x^k) = \{x \mid f(x)'(x - x^k) \leq 0\}$. In some special cases of the EGGF (such as the ellipsoid method [37], [39]), it is important that the sets K^k employed in the algorithm, preserve some properties (for example, convexity and/or connectivity). In these cases notice that if at each step of the EGGF we use a cut determined through a surface $\bar{S}_g = \{x \mid g(x) = 0\}$, where g is a quasiconvex function on K, then the new set $K \cap S_g$, where $S_g = \{x \mid g(x) \leq 0\}$, remains convex. This leads us to conclude that quasiconvex cuts preserve the convexity or the connectivity of the feasible region. In particular, when the problem function f is strongly monotone we can consider quadratic cuts in the EGGF. An example includes a cut where the set lying below the quadratic surface is determined by $S\left(x^k\right) = \left\{x \mid f\left(x^k\right)'\left(x - x^k\right) + \alpha\|x - x^k\|^2 \leq 0\right\}$. Lüthi and Büeler [38] have considered quadratic cuts of this type.

B.2 The convergence of the EGGF

The analysis of the framework relies on the observation that we can measure the "closeness" of a point $x \in K$ to a VIP (or a $WVIP$) solution using some function G. Examples of such functions include the primal or the dual gap function we described in Section 2.2. For a variational inequality problem with a symmetric Jacobian matrix ∇f, an alternate choice could be the corresponding objective function F (that is, when $\nabla F = f$). In what follows, we examine how the properties of the level sets of such a function G imply the convergence of the EGGF. In particular, the following assumptions summarize the key properties.

Assumption 17 *Let X^* be the VIP (or, depending on the context, the $WVIP$) solution set. There exists a function $G : K \rightarrow \mathbb{R}^+$ such that $x^* \in \arg\min_{x \in K} G\left(x\right)$ if and only if $x^* \in X^*$.*

We also assume that the set $S(x)$ lying below the cutting surface at point x, contains the level set $L_\alpha = \{z \mid G\left(z\right) \leq \alpha\}$, for some $\alpha > 0$, i.e., for some $\alpha > 0$, $S(x) \supseteq L_\alpha$.

Examples of sets $S(x)$, include the half space $S(x) = H(x) = \{z \mid a(x)'z \leq b(x)\}$ or the set

$S(x) = \{z \mid a(x)'z + z'Q(x)z \leq b(x)\}$, where $Q(x)$ is a positive semi-definite matrix. Notice that in both examples, set $S(x)$ is a convex set.

Assumption 18 *Given a point* $y \in K$*, and a small enough* $\varepsilon > 0$*, such that* $\min_{x \in X^*} \|x - y\| \leq \varepsilon$*, it follows that* $G(y) \leq c(\varepsilon)$*, where* $\lim_{\varepsilon \to 0} c(\varepsilon) = 0$.

Notice that the previous assumption seems to imply that function G is locally continuous close to the solution set.

In what follows, we establish the convergence of the EGGF when the variational inequality problem satisfies Assumptions
refA1,17, 18.

Theorem 13 *Consider the sequence* $\{x^k\}$ *induced by the EGGF. At each iteration we introduce a cut through a set* $S(x^k)$*. Under Assumptions 1,17,18, every limit point of the sequence* $\{x^k\}$ *is a* VIP *solution.*

Proof. We first examine the properties of the sequences $\{x^k\}$ and $\{K^k\}$. Assumptions 1,17,18 imply that at the kth iteration, the framework adds a cut $S^k = \mathbf{S}(\mathbf{x^k})$, such that for some α^k, $\{z \in K \mid G(z) \leq \alpha^k\} \subseteq S^k$. We note that Assumption 17 implies that the solution set $X^* \subseteq \{z \in K \mid G(z) \leq \alpha^k\} \subseteq S^k$. Since we cut through x^k, it follows that $P^k \not\subseteq S^k$. Moreover, the description of the algorithm implies that $\text{Vol}(K^k) \to 0$, $x^* \in K^k$ and $\text{Vol}(K^k) > 0$, for all k. Let $\bar{K}$ denote the set $\lim_{k \to \infty} K^k$ and $\bar{K}^c$, its complement. Every limit point of the sequence $\{x^k\}$ belongs to the set $\bar{K}$.

Assumption 17 implies that $X^* \subseteq \bar{K}$. Therefore we only need to show that every point $\bar{x} \in \bar{K}$ is a VIP solution. Since $\text{Vol}(K^k) \to 0$ and $\text{Vol}(K^k) > 0$, there exists a point in $\bar{K}^c \cap K$ that is arbitrarily close to some VIP solution x^*. According to Assumption 18, for small enough $\varepsilon > 0$, we can choose this point to be $y^\varepsilon \in \bar{K}^c$ such that $\|y^\varepsilon - x^*\| \leq \varepsilon$ and $G(y^\varepsilon) < c(\varepsilon)$, where $x^* \in X^*$. Since $y^\varepsilon \in \bar{K}^c$, it follows that $G(\bar{x}) < G(y^\varepsilon) \leq c(\varepsilon)$. By letting ε go to zero, we conclude that $G(\bar{x}) = 0$. Consequently, Assumption 17 implies that $\bar{x}$ is a VIP solution. This further implies that every point in the limiting set $\bar{K}$ is a VIP solution. Therefore, we conclude that $\bar{K}$ is the solution set of the VIP. ■

Theorem 14 *Suppose that a* $WVIP$ *satisfies Assumptions 1, 6. Let* $\{x^k\} \subset K$ *be a* *sequence, such that $x^k \in K^k$, where $K^k = K^{k-1} \cap Cut(y^k, x^k)$ and $K^0 = K$. Every limit point of $\{x^k\}$ is a WVIP solution.*

Proof. The dual gap function $C_d(x)$ is the function we will employ in order to measure the closeness of point $x \in K$ to the solution set of the $WVIP$. We will show that for this function, Assumptions 17 and 18 hold. Then the result follows from Theorem 13.

In this theorem, we denote by X^* the $WVIP$ solution set. Since set X^* coincides with set $\{z \mid C_d(z) = 0\}$, in order to prove that Assumption 17 is valid for

Appendix C

Another VIP reformulation

We present this reformulation in a static case. We start by presenting Karush-Kuhn-Tucker (KKT) conditions for the retailer's problem: $\max_{\mathbf{q}\in[0,\mathbf{K}]} \pi_r(\mathbf{w}, \mathbf{q})$.

$$\begin{aligned} w_i - R'(q) + \mu_i - \nu_i &= 0 \\ \mu_i(q_i - K_i) &= 0 \\ \nu_i q_i &= 0 \\ \mu_i, \nu_i \geq 0,\ q_i &\in [0, K_i] \end{aligned}$$

From these conditions it follows that there exists $w = R'(q)$ such that $w_i > w$ implies $q_i = 0$ and $w_i < w$ implies $q_i = K_i$. Using Assumption 16, we obtain that supplier i's problem can be expressed as

$$\max_{w_i}\{q_i(w_i - s_i) \,|\, q_i(w_i - R(q)) = 0,\ w_i \geq 0\},$$

a problem in which feasible set depends on retailer's strategy $\mathbf{q}$. The KKT of all supplier problems can be written together as the following system.

$$\begin{aligned} -q_i + \zeta_i q_i - \xi_i &= 0 \\ \zeta_i q_i(w_i - R'(q)) = 0,\ q_i(w_i - R'(q)) &= 0 \\ \xi_i(w_i - s_i) = 0,\ \xi_i \geq 0,\ w_i &\geq 0 \end{aligned}$$

Concatenating KKTs for all suppliers and the retailer we obtain:

$$w_i - R'(q) + \mu_i - \nu_i = 0$$
$$\mu_i(q_i - K_i) = 0$$
$$\nu_i q_i = 0$$
$$\mu_i, \nu_i \geq 0\, q_i \in [0, K_i]$$
$$-q_i + \zeta_i q_i - \xi_i = 0,\ \zeta_i q_i(w_i - R'(q)) = 0,\ q_i(w_i - R'(q)) = 0$$
$$\xi_i(w_i - s_i) = 0, w_i \geq s_i, \xi_i \geq 0$$

We next suggest the following VIP function: $F^I = \begin{pmatrix} \mathbf{w} - R'(q)\mathbf{e} \\ -\mathbf{Q}(\mathbf{e} - \zeta) \\ \mathbf{Q}(\mathbf{w} - R'(q)\mathbf{e}) \end{pmatrix}$, and the feasible set $K^I = [\mathbf{0}, \mathbf{K}] \times [\mathbf{s}, \infty) \times \mathbb{R}^{n+}$.

We claim that for $z = (\mathbf{q}, \mathbf{w}, \zeta)$, z being solution of VIP $F^I(z^*)'(z - z^*) \geq 0,\ z \in K^I$ is equivalent to $(\mathbf{q}, \mathbf{w})$ being an equilibrium.

Indeed, notice that, $q_i(w_i - R'(q)) = 0$ is a linear constraint with respect to w_i. Moreover, retailer's and suppliers objective's are convex with respect to their respective variables ($\mathbf{q}$ and w_i). Hence given some constraint qualification holds, the VIP is equivalent to GNEP. ([50])

Assumption 19 *$R(q)$ is a concave function.*

Theorem 15 *Given that an equilibrium in pure strategies exists and Assumption 19 holds, a solution of the VIP above constitutes suppliers-retailer game equilibrium.*