END

FILMED

MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

CSL COORDINATED SCIENCE LABORATORY

DECISION AND CONTROL LABORATORY

ADAPTIVE INCENTIVE CONTROLS FOR STACKELBERG GAMES WITH UNKNOWN COST FUNCTIONALS

DTIC
ELECTE
S NOV 2 6 1985 D
B

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

AD-A161 885

REPORT DOCUMENTATION PAGE

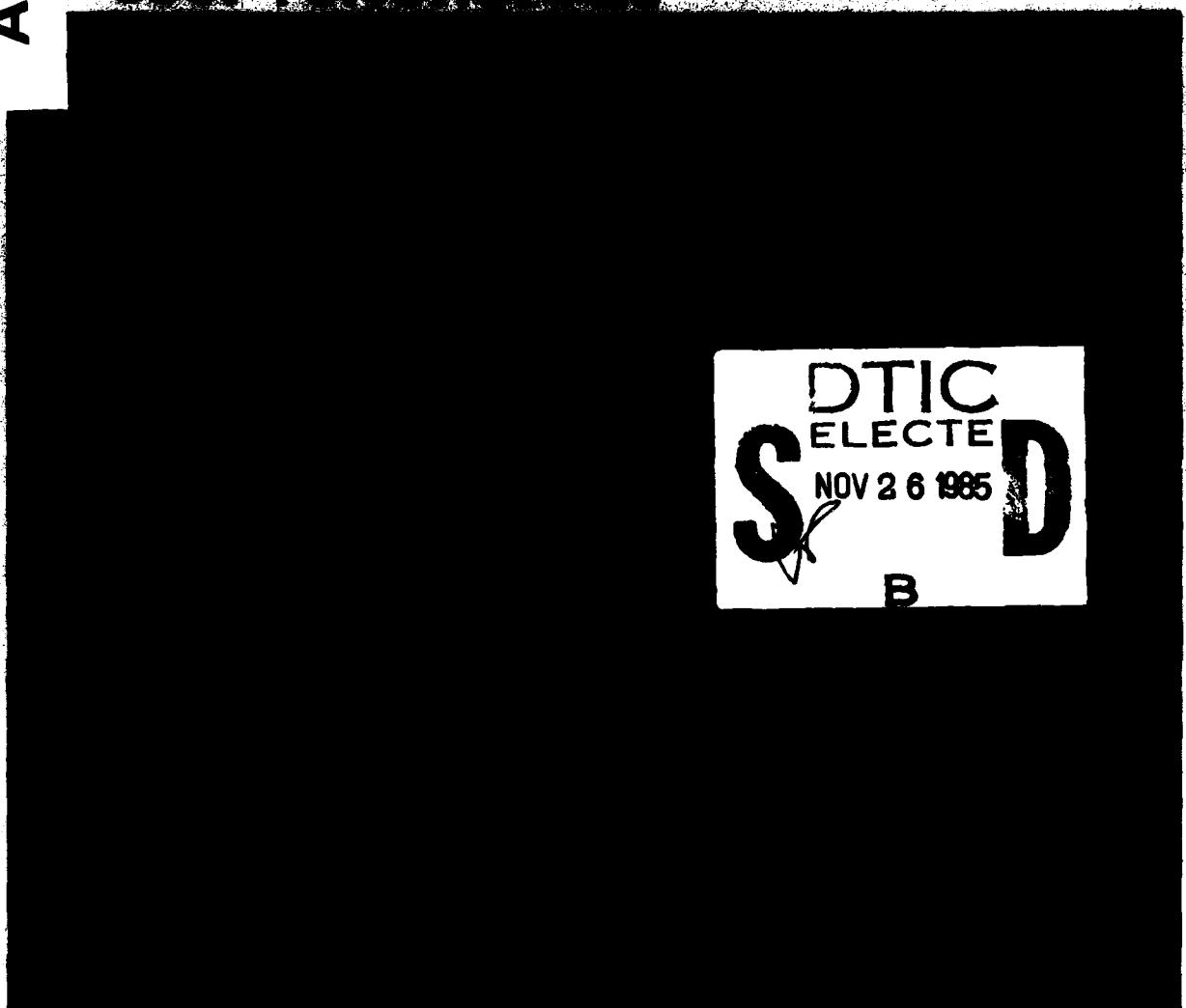| 1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED		1b. RESTRICTIVE MARKINGS NONE			
2a. SECURITY CLASSIFICATION AUTHORITY N/A		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release, distribution unlimited.			
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A					
4. PERFORMING ORGANIZATION REPORT NUMBER(S) R-1022 UILU-ENG-84-2216 (DC-75)		5. MONITORING ORGANIZATION REPORT NUMBER(S) N/A			
6a. NAME OF PERFORMING ORGANIZATION Coordinated Science Laboratory, Univ. of Illinois	6b. OFFICE SYMBOL (If applicable) N/A	7a. NAME OF MONITORING ORGANIZATION Joint Services Electronics Program			
6c. ADDRESS (City, State and ZIP Code) 1101 W. Springfield Avenue Urbana, Illinois 61801		7b. ADDRESS (City, State and ZIP Code) Research Triangle Park, NC 27709			
8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable) N/A	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-84-C-0149			
8c. ADDRESS (City, State and ZIP Code)		10. SOURCE OF FUNDING NOS.			
		PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
11. TITLE (Include Security Classification) Adaptive Incentive Controls for Stackelberg...		N/A	N/A	N/A	N/A
12. PERSONAL AUTHOR(S) Thomas Leo Ting					

13a. TYPE OF REPORT Technical	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day) January 1984	15. PAGE COUNT 54

16. SUPPLEMENTARY NOTATION N/A

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL NONE

DD FORM 1473, 83 APR   EDITION OF 1 JAN 73 IS OBSOLETE.   UNCLASSIFIED

ADAPTIVE INCENTIVE CONTROLS FOR STACKELBERG GAMES
WITH UNKNOWN COST FUNCTIONALS

BY

THOMAS LEO TING

B.S., University of Illinois, 1982
B.S., University of Illinois, 1982

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1984

Thesis Advisor: J. B. Cruz, Jr.

Urbana, Illinois

ACKNOWLEDGMENT

The author would like to extend his gratitude and appreciation to his advisor, Professor J. B. Cruz, Jr., for his invaluable guidance and support. He would also like to thank Mr. R. A. Milito for many helpful discussions.

Sincere thanks are also given to Ms. Rose Harris for her excellent typing.

DTIC
ELECTE
NOV 2 6 1985
B

Accession For

NTIS GRA&I	✓
DTIC TAB	☐
Unannounced	☐
Justification	PER CALL JC

By
Distribution/
Availability Codes

| Dist | Avail and/or Special |
| A-1 | |

QUALITY
CTED

3

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1. History and Motivation

Most control systems problems can be characterized by the following two basic questions: i) Does the problem consist of a single or multiple decision makers? and ii) Does the system under consideration have known or unknown parameters? Although control theory for single decision maker problems has been well developed [1-5], we note that many complicated single decision maker problems can be reformulated into simpler multiple decision maker problems. This is exemplified by the concept of multimodeling strategies for large scale systems [6]. This concept allows a large scale system to be controlled by multiple decision makers using various simplified models of the system. Each decision maker will attempt to individually optimize his own simplified system, but due to modeling uncertainties, there is no assurance that optimization by each individual decision maker will lead to an optimization of the entire system. Therefore, the problem at hand falls nicely into the general framework of stochastic game theory.

Past research in game theory has concentrated mainly on problems involving systems with known parameters [7-9]. However, game theory involving problems with some uncertainties in the system parameters appears to have widespread applications in power systems, industrial systems (as described above), and in various economic and military fields which warrant its consideration. Consider, for example, a situation where several independent firms are selling similar products in the same consumer market. Each firm is attempting to maximize its profit function, which is related through the market structure to its own production level as well as the production

level of its competitors. Realistically, each firm should have complete

knowledge of its own profit function, but not necessarily those of its

competitors. How should each firm proceed to operate? Motivated by this

problem and related examples, we propose to study in this thesis methods

for solving game problems with some uncertainties in the system parameters.

1.2. General Problem Description

There are two basic types of game problems: Nash games and

Stackelberg or leader-follower games. We will concentrate on a basic two-

player Stackelberg game, which is characterized by the fact that one of the

players, known as the leader, has access to more information than the other

player, known as the follower. At every stage of the game each player is

attempting to minimize his own cost functional. For additional simplifica-

tion we focus on the static case (no plant) so that the two cost functionals

depend solely upon the inputs of the two players. These two restrictions

should not cause any loss of generality because we believe that in the

future our results can be extended to both dynamical game problems and game

problems which involve additional players and levels of hierarchy.

In our problem we assume that each player has complete knowledge of

his own cost functional. In addition, we assume the leader knows the structure

of the follower's cost functional, leaving various weighting parameters as

unknown. For a general Stackelberg game we usually have interdependent cost

functionals, and thus the leader's cost will usually depend upon the

follower's input. Consequently, the leader will attempt to use his superior

position to try and influence the follower to react in a manner which helps

minimize the leader's cost. The leader can use his knowledge of the follower's cost functional structure along with some estimates of the unknown parameters to try and predict the follower's possible reaction. This places the leader in a better position to minimize his cost because he may now incorporate the follower's possible reaction into the optimization procedure used to calculate his own input.

One method of implementing this procedure is through an incentive control structure. In this method, the leader applies an input which is functionally dependent upon the difference between the follower's actual input and the follower's input desired by the leader. Incentive control structures have already been applied to Stackelberg games with known system parameters [9,10]. The methods described in these papers do not provide a unique leader's input structure, but the question of selecting a particular input structure based upon a minimum sensitivity approach has been examined [11]. In this thesis we seek to extend the use of incentive control structures to Stackelberg games with unknown cost functionals.

Our approach to this problem is based upon the concept of certainty equivalence [12] and the general theory of self-tuning regulators [13-16]. The algorithm we have devised to solve this problem is basically an adaptive scheme which uses the output of a parameter estimator to self-tune the leader's incentive control structure. A general block diagram of our algorithm is given in Figure 1.1. Our scheme is iterative in nature and is best suited for "on-line" application. Convergence to the optimal incentive control structure is assured for the scalar case and produces satisfactory results when applied to a higher (second) order example.
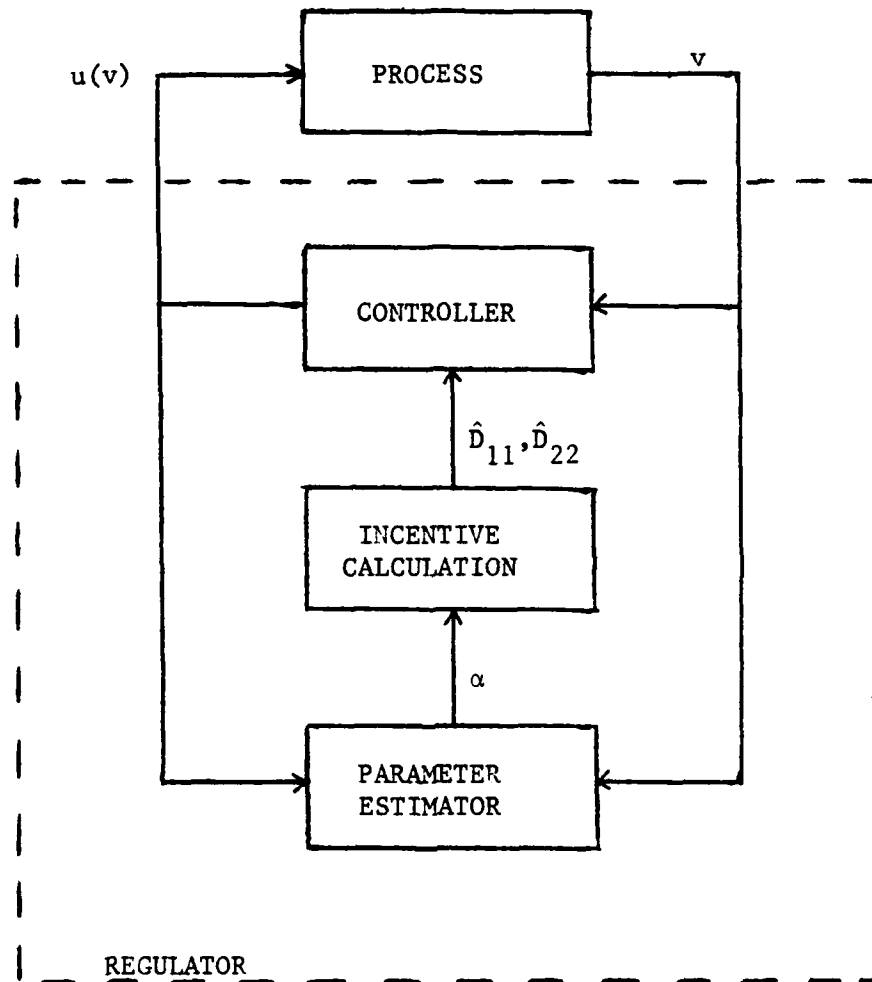
Figure 1.1. A general block diagram of our algorithm.

1.3. Organization of the Thesis

This thesis is organized into five chapters. Chapter 1 is an introduction and a general description of the problem. In Chapter 2 we examine a first order Stackelberg game and develop two methods for iteratively adjusting an incentive control structure. Both of these methods generate controls which converge to the optimal incentive control. Chapter 3 concentrates on developing a corresponding algorithm for a similar problem of higher dimensions. It also describes some additional considerations which are important in developing a higher order incentive control structure. Chapter 4 contains some numerical simulation examples of both the first and second order algorithms and comments on their performance. The thesis is summarized in Chapter 5, which also contains some thoughts on possible areas for further research.

CHAPTER 2

SCALAR STACKELBERG GAMES WITH UNKNOWN COST FUNCTIONALS

2.1. Introduction

In this chapter we study the problem of finding an optimal incentive control for a scalar, static Stackelberg game with unknown cost functionals. We begin by constructing a basic incentive input control structure for the leader. Then we derive an expression for the optimal incentive constant based upon estimates of the unknown system parameters. It is not necessary to estimate the actual values of all of the unknown system parameters since we seek only the information required to obtain the actual optimal incentive constant. The majority of the chapter is devoted to describing two separate parameter estimation schemes, each of which can be used to iteratively produce controls which converge to the optimal incentive control.

2.2. Problem Formulation

Let us consider a scalar, static, two-player Stackelberg game problem where the leader applies the scalar control variable u and the follower applies the scalar control variable v. Each player attempts to minimize for each stage his own quadratic cost functional which are given by

$$J_L = (u-u^t)^2 R_L + (v-v^t)^2 S_L \qquad \text{for the leader} \qquad (2.1a)$$

$$J_F = u^2 R_F + v^2 S_F \qquad \text{for the follower} \qquad (2.1b)$$

where R_L, S_L, R_F, and S_F are positive constants. Clearly J_L and J_F are always ≥ 0 and from the leader's viewpoint (u^t, v^t) is the optimal control pair.

The information structure of this problem is such that each player knows his own cost functional, but the leader also knows the structure of the follower's cost functional. The leader also has the privilege of requiring the follower to play his input v first. The leader then plays his input u and, subsequently, both players may compute their costs for that particular stage. This process is then repeated until an equilibrium condition is reached.

The leader cannot be assured of cooperation from the follower so he will attempt to use his informational advantage to try and force the follower to cooperate. The leader may obtain an estimate (model) of the follower's cost functional by combining his knowledge of the structure of J_F along with estimates for the unknown weighting parameters. The leader may now use this estimate of J_F to simulate the follower's optimization at each stage and thus predict the follower's rational reaction pattern. This allows the leader to select his input u at each stage to optimize his cost functional J_L with respect to both his input u and the follower's possible reactionary input v.

There are many possible methods of implementing this additional information to try and enforce cooperation. We concentrate on an incentive control input structure for the leader. With this structure the leader's input u at each stage is functionally dependent upon the difference between the follower's actual input and the follower's optimal input v^t desired by the leader. The mechanics of the game now proceed as follows [17]:

Step 1. At the beginning of each stage, the leader formulates his incentive control input structure u(v), based on his parameter estimates, and presents it to the follower.

Step 2. The follower calculates his input v by optimizing his cost functional J_F, knowing that the leader's input will be given by the function u(v).

Step 3. The follower plays his input v and the leader's resulting input u is calculated from the input structure u(v). The players then compute their costs for that stage.

At the end of each stage the leader uses the additional information acquired during that stage to update his parameter estimates and then returns to step 1. The game continues in this manner until an equilibrium point is reached. For our purposes, we will define an equilibrium point as an input pair (u_e, v_e) such that each player will play the same input at the next stage. If the leader applies the optimal incentive control input, then obviously the optimal input pair (u^t, v^t) will be the resulting equilibrium point.

2.3. Construction of an Optimal Incentive Control

The leader knows only the structure of J_F and, therefore, he must attempt to estimate the parameters R_F and S_F. Let us denote by $\hat{R}_F^{(i)}$ and $\hat{S}_F^{(i)}$ the leader's estimates of R_F and S_F at stage i. From (2.1b) the leader's estimate of the follower's cost functional at stage i becomes

$$\hat{J}_F^{(i)} = u^2 \hat{R}_F^{(i)} + v^2 \hat{S}_F^{(i)}.$$ (2.2)

Let us collect the unknown system parameters R_F and S_F into a vector denoted by α. Let the vector $\hat{\alpha}^{(i)}$ consist of the estimates $\hat{R}_F^{(i)}$ and $\hat{S}_F^{(i)}$.

At each stage i consider a leader's input of the form

$$u^{(i)} = u^t + D(\hat{\alpha}^{(i)})(v^{(i)} - v^t). \tag{2.3}$$

With this structure the leader's true input $u^{(i)}$ at stage i deviates from his desired optimal input u^t by an incentive constant $D(\hat{\alpha}^{(i)})$ multiplied by the difference between the follower's actual input $v^{(i)}$ at stage i and v^t. Note that when $v^{(i)} = v^t$ we have $u^{(i)} = u^t$ and J_L is minimized. To achieve this minimum the leader must select the appropriate incentive constant $D(\hat{\alpha}^{(i)})$.

Given a particular structure for $u^{(i)}$ as in (2.3), the follower will select an input $v^{(i)}$ which minimizes J_F as given in (2.1b). The leader may simulate this minimization by using his estimate $\hat{J}_F^{(i)}$ given in (2.2) and computing the solution to

$$\frac{d\hat{J}_F^{(i)}}{d\hat{v}^{(i)}} = \frac{\partial \hat{J}_F^{(i)}}{\partial \hat{v}^{(i)}} + \frac{\partial \hat{J}_F^{(i)}}{\partial u^{(i)}} \frac{\partial u^{(i)}}{\partial \hat{v}^{(i)}} = 0 \tag{2.4}$$

where $\hat{v}^{(i)}$ represents the leader's estimate of the follower's input $v^{(i)}$ at stage i. Solving this expression we obtain

$$\frac{d\hat{J}_F^{(i)}}{d\hat{v}^{(i)}} = 2\hat{v}^{(i)}\hat{S}_F^{(i)} + 2\hat{R}_F^{(i)}D(\hat{\alpha}^{(i)})[u^t + D(\hat{\alpha}^{(i)})(\hat{v}^{(i)} - v^t)] = 0 \tag{2.5}$$

which reduces to

$$\hat{v}^{(i)} = \frac{D(\hat{\alpha}^{(i)})[D(\hat{\alpha}^{(i)})v^t - u^t]}{(D(\hat{\alpha}^{(i)})^2 + \hat{S}_F^{(i)}/\hat{R}_F^{(i)})}. \tag{2.6}$$

The leader wishes to select $D(\hat{\alpha}^{(i)})$ so that $\hat{v}^{(i)} = v^t$ at each stage. Setting $\hat{v}^{(i)} = v^t$ in (2.6) produces

$$D(\hat{\alpha}^{(i)}) = -\frac{\hat{S}_F^{(i)} v^t}{\hat{R}_F^{(i)} u^t} \qquad (2.7)$$

as the optimal incentive constant. We can see that $D(\hat{\alpha}^{(i)})$ does not depend explicitly upon both $\hat{R}_F^{(i)}$ and $\hat{S}_F^{(i)}$, but merely upon their ratio. Thus, the leader need only estimate the ratio

$$\hat{x}^{(i)} = \hat{S}_F^{(i)} / \hat{R}_F^{(i)}. \qquad (2.8)$$

With this notation the expression for the optimal incentive constant in (2.7) reduces to

$$D(\hat{\alpha}^{(i)}) = -\hat{x}^{(i)} \frac{v^t}{u^t}. \qquad (2.9)$$

2.4. Parameter Estimations and Updating Techniques

The leader begins his estimation of $x = S_F/R_F$ by making an initial guess $\hat{x}^{(0)}$. Since S_F and R_F were both positive, the ratio x is always positive and the initial guess $\hat{x}^{(0)}$ can be constrained to positive values. Using $\hat{x}^{(0)}$, the leader computes $D(\hat{x}^{(0)})$ via (2.9) and his input structure $u^{(0)}$ via (2.3), and presents it to the follower. After the follower computes $v^{(0)}$, the true value of $u^{(0)}$ is obtained and both players may compute their costs for step zero. At the end of this stage the leader must have a method for updating $\hat{x}^{(0)}$ which incorporates the new information acquired during stage zero. We now consider the formulation of a general updating procedure to be used following any stage i.

To update $\hat{x}^{(i)}$ we first calculate the follower's actual input $v^{(i)}$ at step i, with $u^{(i)}$ taking the form given in (2.3) and $D(\hat{\alpha}^{(i)})$ as in (2.9). This calculation is similar to the one in equations (2.4) and (2.5) and yields

$$v^{(i)} = \frac{D(\hat{\alpha}^{(i)})[D(\hat{\alpha}^{(i)})v^t - u^t]}{[D(\hat{\alpha}^{(i)})^2 + x]}. \tag{2.10}$$

We would like to update $\hat{x}^{(i)}$ so that $v^{(i)}$ converges to $\hat{v}^{(i)}$. The value of $\hat{v}^{(i)}$ was given in (2.6), and by our selection of $D(\hat{\alpha}^{(i)})$ this reduces to

$$\hat{v}^{(i)} = v^t. \tag{2.11}$$

Therefore, we actually want to update $\hat{x}^{(i)}$ in such a way that $v^{(i)}$ converges to $\hat{v}^{(i)} = v^t$, the optimal follower's input from the leader's viewpoint. We have devised two separate methods for updating $\hat{x}^{(i)}$.

2.4.1. Error function method

Consider the following function of $x^{(i)}$, which measures the error between $v^{(i)}$ and $\hat{v}^{(i)}$

$$E_i(\hat{x}^{(i)}) = \frac{1}{2}(v^{(i)} - \hat{v}^{(i)})^2 = \frac{1}{2}(v^{(i)} - v^t)^2. \tag{2.12}$$

Substituting (2.9) into (2.10) we obtain

$$v^{(i)} = \frac{\hat{x}^{(i)^2} \frac{v^{t^3}}{u^{t^2}} + \hat{x}^{(i)} v^t}{\hat{x}^{(i)^2} \frac{v^{t^2}}{u^{t^2}} + x} + \frac{\hat{x}^{(i)^2} v^{t^3} + x^{(i)} u^{t^2} v^t}{\hat{x}^{(i)^2} v^{t^2} + x u^{t^2}}. \tag{2.13}$$

From (2.13) we can calculate

$$v^{(i)} - v^t = \frac{u^{t^2} v^t (\hat{x}^{(i)} - x)}{\hat{x}^{(i)^2} v^{t^2} + xu^{t^2}}. \tag{2.14}$$

Thus for $u^t \neq 0$ and $v^t \neq 0$ we have $v^{(i)} = v^t$ if and only if $\hat{x}^{(i)} = x$. We also find by differentiating (2.14) and treating x as a constant

$$\frac{\partial v^{(i)}}{\partial \hat{x}^{(i)}} = \frac{(2\hat{x}^{(i)} v^{t^3} + u^{t^2} v^t)(\hat{x}^{(i)^2} v^{t^2} + xu^{t^2}) - 2\hat{x}^{(i)} v^{t^2}(\hat{x}^{(i)^2} v^{t^3} + \hat{x}^{(i)} u^{t^2} v^t)}{(\hat{x}^{(i)^2} v^{t^2} + x u^{t^2})^2}. \tag{2.15}$$

Let us restrict the domain of $E_i(\hat{x}^{(i)})$ to positive values of $\hat{x}^{(i)}$. Note that $E_i(\hat{x}^{(i)})$ as given in (2.12) is a positive definite function because $E_i(\hat{x}^{(i)}) = 0$ only when $\hat{x}^{(i)} = x$, and $E_i(\hat{x}^{(i)}) > 0$ at all other values of $\hat{x}^{(i)}$. Our goal of adjusting $\hat{x}^{(i)}$ such that $v^{(i)}$ converges to v^t can now be accomplished by adjusting $\hat{x}^{(i)}$ such that $E_i(\hat{x}^{(i)})$ converges to zero. We propose to update $\hat{x}^{(i)}$ by using the following gradient technique on $E_i(\hat{x}^{(i)})$.

$$\hat{x}^{(i+1)} = \hat{x}^{(i)} - \gamma \nabla_{\hat{x}^{(i)}} E_i = \hat{x}^{(i)} - \gamma(v^{(i)} - v^t)\nabla_{\hat{x}^{(i)}} v^{(i)}. \tag{2.16}$$

At the end of each stage i, the leader knows the value of the input $v^{(i)}$ which the follower has just applied, and can easily calculate the quantity $v^{(i)} - v^t$. In addition, $\nabla_{\hat{x}^{(i)}} v^{(i)}$ can be calculated from (2.15) by using the known quantities v^t and u^t, the current estimate $\hat{x}^{(i)}$, and by substituting $\hat{x}^{(i)}$ as an estimate for the unknown constant x. Thus, the leader has enough information to implement the updating procedure (2.16). We know that x is always positive so it is reasonable to require that $\hat{x}^{(i)}$ always remain positive. This can be accomplished by an appropriate choice of the step size γ. Our updating procedure leads to the following theorem.

__Theorem 1__: The gradient technique for updating $\hat{x}^{(i)}$ based upon the error function $E_i(\hat{x}^{(i)})$, applied with a sufficiently small step size γ and by using $\hat{x}^{(i)}$ as an estimate for x in $\nabla_{\hat{x}^{(i)}} v^{(i)}$, produces estimates $\hat{x}^{(i)}$ which converge to the actual value of x.

__Proof__: The gradient algorithm for updating $\hat{x}^{(i)}$ is given by equation (2.16).

Substituting (2.14) and (2.15) into (2.16) we obtain the updating equation

$$\hat{x}^{(i+1)} = \hat{x}^{(i)} - \gamma \left[\frac{R_F v^t \hat{x}^{(i)} - S_F v^t}{S_F + R_F \hat{x}^{(i)2} \frac{v^{t2}}{u^{t2}}} \right] \left[\frac{u^{t4} v^t x - \hat{x}^{(i)2} u^{t2} v^{t3} + 2x\hat{x}^{(i)} u^{t2} v^{t3}}{(x u^{t2} + \hat{x}^{(i)2} v^{t2})^2} \right] . \quad (2.17)$$

The actual value of x is unknown to the leader so we substitute our estimate $\hat{x}^{(i)}$ for x in (2.17) to produce

$$\hat{x}^{(i+1)} = \hat{x}^{(i)} - \gamma \left[\frac{R_F v^t \hat{x}^{(i)} - S_F v^t}{S_F + R_F \hat{x}^{(i)2} \frac{v^{t2}}{u^{t2}}} \right] \left[\frac{u^{t4} v^t \hat{x}^{(i)} + \hat{x}^{(i)2} u^{t2} v^{t3}}{(\hat{x}^{(i)} u^{t2} + \hat{x}^{(i)2} v^{t2})^2} \right] . \quad (2.18)$$

From (2.18) we see that the only equilibrium point for our updating scheme (if $v^t, u^t \neq 0$) occurs when

$$R_F v^t \hat{x}^{(i)} = S_F v^t \implies \hat{x}^{(i)} = \frac{S_F}{R_F} = x. \quad (2.19)$$

From (2.19) we see that this updating method has the desired equilibrium point.

To prove that $\hat{x}^{(i)}$ converges to x we consider two separate cases:

i) Case 1: If $\hat{x}^{(i)} > x$.

ii) Case 2: If $\hat{x}^{(i)} < x$.

For simplicity we rewrite (2.18) as

$$\hat{x}^{(i+1)} = \hat{x}^{(i)} - \gamma \Delta x \quad \text{where} \quad \Delta x = \left[\frac{R_F v^{t} \hat{x}^{(i)} - S_F v^{t}}{S_F + R_F \hat{x}^{(i)^2} \frac{v^{t^2}}{u^{t^2}}} \right] \left[\frac{u^{t^4} x^{t} \hat{x}^{(i)} + \hat{x}^{(i)^2} u^{t^2} v^{t^3}}{(\hat{x}^{(i)} u^{t^2} + \hat{x}^{(i)^2} v^{t^2})^2} \right].$$

$$(2.20)$$

By requiring $\hat{x}^{(i)}$ to be positive at all times, and since R_F and S_F are positive

constants, we see that the denominator of Δx is always positive. Therefore,

we concentrate our interest on the numerator of Δx because sign Δx = sign Δx_{NUM}.

From (2.20) we have

$$\Delta x_{NUM} = (v^{t^2} u^{t^2} R_F)(\hat{x}^{(i)} - x)(u^{t^2} \hat{x}^{(i)} + v^{t^2} \hat{x}^{(i)^2}). \qquad (2.21)$$

Examination of the case $\hat{x}^{(i)} > x$

In this case we want to show that $\Delta x > 0$. It is obvious that the

first two terms in (2.21) are positive because R_F is a positive constant

and $\hat{x}^{(i)} - x$ is also positive. Furthermore, the third term of (2.21) is also

positive by our previous requirement that $\hat{x}^{(i)}$ be positive at all times.

The product of three positive quantities is always positive so (2.21) implies

that Δx is positive whenever $\hat{x}^{(i)}$ is greater than x. Thus, whenever $\hat{x}^{(i)}$ is

greater than x, our updating procedure (2.20) will produce an updated

estimate $\hat{x}^{(i+1)}$ which is less than $\hat{x}^{(i)}$ as desired.

Examination of the case $\hat{x}^{(i)} < x$

In this case we want to show that $\Delta x < 0$. Similarly as in the

previous case we find that the first and third terms of (2.21) are positive.

However, since $\hat{x}^{(i)}$ is less than x, the second term in (2.21), $\hat{x}^{(i)} - x$, is

negative. This makes the expression Δx in (2.21) negative whenever $\hat{x}^{(i)}$ is

less than x. Therefore, if $\hat{x}^{(i)} < x$ our updating procedure (2.20) will produce an updated estimate $\hat{x}^{(i+1)}$ which is greater than $\hat{x}^{(i)}$ as desired.

We have just shown that if $\hat{x}^{(i)}$ is greater than (less than) x then our updating procedure (2.20) produces a next estimate $\hat{x}^{(i+1)}$ which is less than (greater than) $\hat{x}^{(i)}$. By choosing γ small enough such that $\hat{x}^{(i+1)} - x$ has the same sign as $\hat{x}^{(i)} - x$, we insure that the differences $|\hat{x}^{(i)} - x|$ form a monotonically decreasing sequence of errors which converge to its lower bound of zero. Therefore, for a properly chosen constant γ the updating scheme (2.20) produces estimates $\hat{x}^{(i)}$ which converge to x. This completes the proof.

This updating scheme produces estimates $\hat{x}^{(i)}$ which display a one-sided convergence to x, with the direction of convergence depending upon the selection of $\hat{x}^{(0)}$. If $\hat{x}^{(0)} > x$ then $\hat{x}^{(i)}$ converges to x from above and if $\hat{x}^{(0)} < x$ then $\hat{x}^{(i)}$ converges to x from below. Note that when $\hat{x}^{(i)}$ converges to x, $D(\hat{a}^{(i)})$ in (2.7) converges to the optimal incentive constant

$$D^* = - \frac{S_F v^t}{R_F u^t} \ . \tag{2.22}$$

A leader's input structure of the form (2.3) using (2.22) will generate the optimal input pair (u^t, v^t).

2.4.2. Gradient on J_L method

A second method of updating $\hat{x}^{(i)}$ uses a gradient technique which is based upon minimizing J_L. Intuitively this method is the most logical one because it adjusts $\hat{x}^{(i)}$ in a manner which creates the greatest decrease in the value of J_L, which is precisely the overall goal. This updating

scheme is given by

$$\hat{x}^{(i+1)} = \hat{x}^{(i)} - \gamma \nabla_{\hat{x}^{(i)}} J_L. \qquad (2.23)$$

Using (2.1a), (2.3) and (2.9) we can rewrite J_L as

$$J_L = D(\hat{\alpha}^{(i)})^2 R_L (v^{(i)} - v^t)^2 + S_L (v^{(i)} - v^t)^2 = \left(\hat{x}^{(i)^2} \frac{v^{t^2}}{u^{t^2}} R_L + S_L \right)(v^{(i)} - v^t)^2. \qquad (2.24)$$

We must be careful when calculating $\nabla_{\hat{x}^{(i)}} J_L$ to include the dependence of both $D(\hat{\alpha}^{(i)})$ and $v^{(i)}$ upon $\hat{x}^{(i)}$. By the chain rule we have

$$\nabla_{\hat{x}^{(i)}} J_L = \frac{dJ_L}{d\hat{x}^{(i)}} = \frac{\partial J_L}{\partial v^{(i)}} \frac{\partial v^{(i)}}{\partial \hat{x}^{(i)}} + \frac{\partial J_L}{D(\hat{\alpha}^{(i)})} \frac{D(\hat{\alpha}^{(i)})}{\partial \hat{x}^{(i)}}. \qquad (2.25)$$

From (2.9) and (2.24) we can find

$$\frac{\partial D(\hat{\alpha}^{(i)})}{\partial \hat{x}^{(i)}} = -v^t/u^t \qquad (2.26)$$

$$\frac{\partial J_L}{\partial D(\hat{\alpha}^{(i)})} = 2D(\hat{\alpha}^{(i)}) R_L (v^{(i)} - v^t)^2 = -2\hat{x}^{(i)} \frac{v^t}{u^t} R_L (v^{(i)} - v^t)^2 \qquad (2.27)$$

$$\frac{\partial J_L}{\partial v^{(i)}} = 2(v^{(i)} - v^t)[D(\hat{\alpha}^{(i)})^2 R_L + S_L] = 2(v^{(i)} - v^t)\left[\frac{\hat{x}^{(i)^2} v^{t^2}}{u^{t^2}} R_L + S_L \right]. \qquad (2.28)$$

Substituting (2.26)-(2.28) into (2.25) we can write

$$\nabla_{\hat{x}^{(i)}} J_L = 2\hat{x}^{(i)} R_L \frac{v^{t^2}}{u^{t^2}} (v^{(i)} - v^t)^2 + 2(v^{(i)} - v^t)\left[\frac{\hat{x}^{(i)^2} v^{t^2}}{u^{t^2}} R_L + S_L \right] \frac{\partial v^{(i)}}{\partial \hat{x}^{(i)}}. \qquad (2.29)$$

At the end of each stage i, the values of the inputs $v^{(i)}$ and $u^{(i)}$ are known so all of the quantities in (2.26)-(2.28) can easily be calculated

by the leader. The quantity $\dfrac{\partial v^{(i)}}{\partial x^{(i)}}$ is found by differentiating (2.13) as

before, and treating x as a constant. The resulting expression is the same

as in (2.15). Once again since the actual value of x is unknown to the

leader, he replaces x in (2.15) by his estimate $\hat{x}^{(i)}$ which yields

$$\frac{\partial v^{(i)}}{\partial \hat{x}^{(i)}} = \frac{u^{t^4}v^t\hat{x}^{(i)} + \hat{x}^{(i)^2}u^{t^2}v^{t^3}}{(\hat{x}^{(i)^2}v^{t^2} + \hat{x}^{(i)}u^{t^2})^2}. \tag{2.30}$$

Using this expression for $\dfrac{\partial v^{(i)}}{\partial \hat{x}^{(i)}}$ and the expression for $v^{(i)} - v^t$ in (2.14), we

write (2.29) as

$$\nabla_{\hat{x}^{(i)}} J_L = 2\hat{x}^{(i)}R_L \frac{v^{t^2}}{u^{t^2}}\left(\frac{u^{t^2}v^t(\hat{x}^{(i)}-x)}{\hat{x}^{(i)^2}v^{t^2}+xu^{t^2}}\right)^2$$

$$+ 2\left(\frac{\hat{x}^{(i)^2}v^{t^2}}{u^{t^2}}R_L+S_L\right)\left(\frac{u^{t^2}v^t(\hat{x}^{(i)}-x)}{\hat{x}^{(i)^2}v^{t^2}+xu^{t^2}}\right)\left(\frac{u^{t^4}v^t\hat{x}^{(i)}+\hat{x}^{(i)^2}u^{t^2}v^{t^3}}{(\hat{x}^{(i)^2}v^{t^2}+\hat{x}^{(i)}u^{t^2})^2}\right)$$

$$= [2(\hat{x}^{(i)}-x)u^{t^2}v^{t^2}(u^{t^2}\hat{x}^{(i)}+\hat{x}^{(i)^2}v^{t^2})]$$

$$\times \frac{[R_Lv^{t^2}(\hat{x}^{(i)}-x)(\hat{x}^{(i)^2}v^{t^2}+\hat{x}^{(i)}u^{t^2})+(\hat{x}^{(i)^2}v^{t^2}R_L+S_Lu^{t^2})(\hat{x}^{(i)^2}v^{t^2}+xu^{t^2})]}{(\hat{x}^{(i)^2}v^{t^2}+xu^{t^2})^2(\hat{x}^{(i)^2}v^{t^2}+\hat{x}^{(i)}u^{t^2})^2}.$$

$$\tag{2.31}$$

In the updating equation (2.23) let us again choose γ sufficiently small such

that $\hat{x}^{(i)}$ remains positive at all stages of the iteration. This leads to

the following theorem.

Theorem 2: For an initial guess $\hat{x}^{(0)}$ large enough such that $\hat{x}^{(0)} > x$, the

iterative scheme detailed in (2.23), applied by using an appropriately small

step size γ and by substituting $\hat{x}^{(i)}$ as an estimate for x in the expression $\partial v^{(i)}/\partial \hat{x}^{(i)}$, produces estimates $\hat{x}^{(i)}$ which converge to the true value of x.

<u>Proof</u>: The algorithm for updating $\hat{x}^{(i)}$, given by (2.23), uses $\nabla_{\hat{x}^{(i)}} J_L$, which takes the form of (2.31).

From (2.31) we see that our updating scheme has equilibrium points (if v^t, $u^t \neq 0$) when

$$\hat{x}^{(i)} = x - \frac{(\hat{x}^{(i)^2} v^{t^2} + xu^{t^2})^2}{v^{t^2}(\hat{x}^{(i)^2} v^{t^2} + \hat{x}^{(i)} u^{t^2})} \tag{2.32a}$$

$$\hat{x}^{(i)} = x. \tag{2.32b}$$

Obviously, (2.32b) is the desired equilibrium point while (2.32a) is an undesirable equilibrium point which we seek to avoid. By choosing γ small enough such that $\hat{x}^{(i)} - x \geq 0$ for all i, we can disregard the equilibrium point (2.32a) because it is smaller than x.

It is obvious that the denominator of $\nabla_{\hat{x}^{(i)}} J_L$ in (2.31) is positive, so we concentrate our attention on the numerator because sign $\nabla_{\hat{x}^{(i)}} J_L$ = sign $\nabla_{\hat{x}^{(i)}} J_{L_{NUM}}$. From (2.31) we have

$$\nabla_{\hat{x}^{(i)}} J_{L_{NUM}} = [2u^{t^2} v^{t^2}(u^{t^2} \hat{x}^{(i)} + \hat{x}^{(i)^2} v^{t^2})(\hat{x}^{(i)} - x)][R_L v^{t^2}(\hat{x}^{(i)} - x)(\hat{x}^{(i)^2} v^{t^2}$$
$$+ \hat{x}^{(i)} u^{t^2}) + (\hat{x}^{(i)^2} v^{t^2} R_L + S_L u^{t^2})(\hat{x}^{(i)^2} v^{t^2} + xu^{t^2})]. \tag{2.33}$$

We want to show that $\nabla_{\hat{x}^{(i)}} J_{L_{NUM}}$ in (2.33) is positive. The difference $\hat{x}^{(i)} - x$ is always greater than zero by our choice of γ and $\hat{x}^{(0)}$, so by inspection, we see that all of the terms in (2.33) are positive. This makes both $\nabla_{\hat{x}^{(i)}} J_L$ and $\nabla_{\hat{x}^{(i)}} J_{L_{NUM}}$ positive. Therefore, whenever $\hat{x}^{(i)} > x$ our

updating scheme (2.23) produces an updated estimate $\hat{x}^{(i+1)}$ such that $\hat{x}^{(i+1)} < \hat{x}^{(i)}$ as desired. This completes the proof.

Unfortunately, if $\hat{x}^{(0)}$ is chosen such that $\hat{x}^{(0)} < x$, then we can no longer guarantee that the estimates $\hat{x}^{(i)}$ produced by the algorithm in (2.23) converge to the actual value of x. We can, however, prove the following theorem.

Theorem 3: For an initial guess $\hat{x}^{(0)} < x$, the iterative scheme of (2.33), applied with an appropriately small step size γ and by substituting $\hat{x}^{(i)}$ as an estimate for x in the expression $\partial v^{(i)}/\partial \hat{x}^{(i)}$, produces estimates $\hat{x}^{(i)}$ which converge to the true value of x provided $\hat{x}^{(0)}$ satisfies the following conditions

$$\hat{x}^{(0)} - x = F(\hat{x}^{(0)}) > - \frac{(\hat{x}^{(0)^2} v^{t^2} + xu^{t^2})^2}{v^{t^2}(\hat{x}^{(0)^2} v^{t^2} + \hat{x}^{(0)} u^{t^2})} \qquad (2.34)$$

$$\frac{\partial F(\hat{x}^{(0)})}{\hat{x}^{(0)}} = \frac{x^2 u^{t^6} + 2x^2 \hat{x}^{(0)} v^{t^2} u^{t^4} - 3\hat{x}^{(0)^4} v^{t^2} u^{t^4} - 2\hat{x}^{(0)^5} v^{t^6} - 2\hat{x}^{(0)^2} xv^{t^2} u^{t^4}}{v^{t^4}(\hat{x}^{(0)^2} v^{t^2} + \hat{x}^{(0)} u^{t^2})^2} < 0. \qquad (2.35)$$

Proof: Once again the updating algorithm (2.23) uses the expression $\nabla_{\hat{x}^{(i)}} J_L$ given in (2.31). Consequently, this updating scheme has the same two equilibrium points given in (2.32a,b). The denominator of $\nabla_{\hat{x}^{(i)}} J_L$ in (2.31) is positive by inspection so sign $\nabla_{\hat{x}^{(i)}} J_L$ = sign $\nabla_{\hat{x}^{(i)}} J_{L_{NUM}}$. This allows us to concentrate on the reduced expression in (2.33).

We want to show that $\nabla_{\hat{x}^{(i)}} J_{L_{NUM}}$ in (2.33) is negative. However, this situation is more complicated than in Theorem 2 because we must also avoid the undesirable equilibrium point given in (2.32a). We can insure that the estimates $\hat{x}^{(i)}$ are always smaller than x by an appropriate choice

of γ. This restricts the first term in (2.33) to negative values. Therefore, a necessary condition for $\nabla_{\hat{x}^{(i)}} J_{L_{NUM}}$ to be negative is that the second term in (2.33) be positive.

Our necessary condition is

$$_v t^2 (\hat{x}^{(i)} - x)(\hat{x}^{(i)2} _v t^2 + \hat{x}^{(i)} _u t^2) + (\hat{x}^{(i)2} _v t^2 + x_u t^2)^2 > 0 \qquad (2.36)$$

which reduces to

$$\hat{x}^{(i)} > x - \frac{(\hat{x}^{(i)2} _v t^2 + x_u t^2)^2}{_v t^2 (\hat{x}^{(i)2} _v t^2 + \hat{x}^{(i)} _u t^2)} . \qquad (2.37)$$

The expression on the right hand side of (2.37) is precisely the value of the undesirable equilibrium point given in (2.32a). We see that if (2.37) holds then $\nabla_{\hat{x}^{(i)}} J_{L_{NUM}} < 0$ and the updating scheme in (2.23) will produce an $\hat{x}^{(i+1)}$ which is greater than $\hat{x}^{(i)}$ as desired. We now seek to find conditions on $\hat{x}^{(0)}$ to insure that $\hat{x}^{(i)}$ converges to x.

Rewrite (2.37) as

$$\hat{x}^{(i)} - x > - \frac{(\hat{x}^{(i)2} _v t^2 + x_u t^2)^2}{_v t^2 (\hat{x}^{(i)2} _v t^2 + \hat{x}^{(i)} _u t^2)} = F(\hat{x}^{(i)}) . \qquad (2.38)$$

Now differentiate (2.38) with respect to $\hat{x}^{(i)}$ to obtain

$$\frac{\partial F(\hat{x}^{(i)})}{\partial x^{(i)}} = \frac{x^2 _u t^6 + 2x^2 \hat{x}^{(i)} _v t^2 _u t^4 - 3\hat{x}^{(i)4} _u t^2 _v t^4 - 2\hat{x}^{(i)5} _v t^6 - 2\hat{x}^{(i)2} _{xv} t^2 _u t^4}{_v t^4 (\hat{x}^{(i)2} _v t^2 + \hat{x}^{(i)} _u t^2)^2} . (2.39)$$

Notice that the denominator of this equation is always positive and that as $\hat{x}^{(i)}$ increases, the numerator decreases. Consequently, if $\frac{\partial F(\hat{x}^{(i)})}{\partial x^{(i)}}$ is negative, and $\hat{x}^{(i)} < \hat{x}^{(i+1)}$, then $\frac{\partial F(\hat{x}^{(i+1)})}{\partial x^{(i+1)}}$ must also be

negative. Thus for a negative value of $\dfrac{\partial F(\hat{x}^{(i)})}{\partial \hat{x}^{(i)}}$ we can conclude that if

(2.38) holds for $\hat{x}^{(i)}$ then it must also hold for $\hat{x}^{(i+1)}$ since

$$\hat{x}^{(i+1)} - x > \hat{x}^{(i)} - x \quad \text{and} \quad F(\hat{x}^{(i+1)}) < F(\hat{x}^{(i)}). \qquad (2.40)$$

Combining these results, we can guarantee that whenever $\hat{x}^{(i)} < x$ our

algorithm (2.23), applied as stated in Theorem 3, will converge to x if

both equation (2.38) holds and $\dfrac{\partial F(\hat{x}^{(i)})}{\partial \hat{x}^{(i)}}$, as given in (2.39), is negative.

If we select γ small enough such that $\hat{x}^{(i)}-x$ and $\hat{x}^{(i+1)}-x$ have the same

sign, then we may generalize the above conditions to $\hat{x}^{(0)}$. Therefore,

if (2.38) holds for $\hat{x}^{(0)}$ and (2.39) is negative for $\hat{x}^{(0)}$, then our subsequent

estimates $\hat{x}^{(i)}$ will converge to x. These are precisely the sufficiency

conditions given in the statement of the theorem and thus the theorem is

proved.

It is interesting to note that the sufficiency condition (2.34)

of Theorem 3 explicitly requires that the initial guess be larger than the

undesirable equilibrium point given by (2.32a). In addition, the

sufficiency condition (2.35) which states that $\dfrac{\partial F(\hat{x}^{(0)})}{\partial \hat{x}^{(0)}} < 0$ is somewhat

analogous to insuring a locally convex function for our gradient technique.

However, the leader has no a priori information about the value of x and

thus has no way of knowing whether his initial guess will satisfy the

sufficiency conditions of Theorem 3. Therefore, in applications it is more

logical for the leader to select a somewhat larger value for his initial

guess $\hat{x}^{(0)}$, since Theorem 2 guarantees convergence to the actual value of x

from above. In general, the convergence, if it occurs, will be one-sided

and will depend upon whether $\hat{x}^{(0)}$ is greater than or less than x. If $\hat{x}^{(i)}$

does converge to x we are again assured that $D(\hat{x}^{(i)})$ will converge to the

optimal incentive constant D^* in (2.22). The resulting leader's input structure from (2.3) by using (2.22) will generate the optimal input pair (u^t, v^t).

CHAPTER 3

HIGHER ORDER STACKELBERG GAMES WITH UNKNOWN COST FUNCTIONALS

3.1. Introduction

Chapter 2 developed methods for finding an optimal incentive control for scalar, static, two-player Stackelberg games with unknown cost functionals. We now seek to develop corresponding methods for similar problems of higher dimension. In this chapter we concentrate on a second order game and again adopt a basic incentive input control structure for the leader. In this case we no longer have an incentive constant, but rather a 2×2 incentive matrix which provides great flexibility. The leader is free to select the structure of this incentive matrix, the elements of which depend upon his estimates of the unknown cost functionals. The resulting optimal incentive matrix may or may not be unique, depending upon the structure chosen by the leader. Generally, the leader may use his freedom in selecting the incentive matrix to satisfy specific design considerations as described at the end of the chapter.

Paralleling the scalar case, it is again unnecessary to precisely estimate the actual values of the unknown system parameters. It suffices to obtain enough information about these parameters to find the elements of the optimal incentive matrix. The parameter estimate updating method utilizing a gradient technique on J_L provides good results and its details are described in Section 3.4.

3.2. Problem Formulation

We now consider a general problem similar to that of Chapter 2 but with higher dimensions. Assume that the control inputs are given by the 2×1 vectors

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad \text{for the leader} \quad \text{AND} \quad v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad \text{for the follower.}$$

Each player again tries to minimize at each stage his own quadratic cost functional given by

$$J_L = (u-u^t)'R_L(u-u^t) + (v-v^t)'S_L(v-v^t) \quad \text{for the leader} \tag{3.1a}$$

$$J_F = u'R_F u + v'S_F v \quad \text{for the follower} \tag{3.1b}$$

where R_L, S_L, R_F, and S_F are all symmetric, positive definite 2×2 matrices. Thus, J_F and J_L are always ≥ 0 and from the leader's viewpoint an optimal solution is given by the input control pair

$$u^t = \begin{bmatrix} u_1^t \\ u_2^t \end{bmatrix} \quad \text{AND} \quad v^t = \begin{bmatrix} v_1^t \\ v_2^t \end{bmatrix}.$$

We assume that the information structure of this problem and the steps followed in playing the game are the same as described in Section 2.2. At the end of each stage the leader will use the new information obtained during that stage to update his parameter estimates and the next stage of the game will begin. The players will continue to play until an equilibrium point is reached as previously defined.

3.3. Construction of an Optimal Incentive Control

The weighting matrices R_F and S_F of the follower's cost functional are unknown and thus the leader will attempt to estimate them. Let $\hat{R}_F^{(i)}$ and $\hat{S}_F^{(i)}$ denote the leader's estimates of the matrices R_F and S_F at stage i. There are actually only five unknown elements in the matrices R_F and S_F since they are both symmetric and can be scaled such that $R_{11} = 1$. Let us stack the unknown elements of R_F and S_F into a vector α. Then $\hat{\alpha}^{(i)}$ is the stacked vector representing the estimates $\hat{R}_{F12}^{(i)}$, $\hat{R}_{F22}^{(i)}$, $\hat{S}_{F11}^{(i)}$, $\hat{S}_{F12}^{(i)}$, and $\hat{S}_{F22}^{(i)}$ of these unknown elements at stage i.

Consider an input of the form

$$u^{(i)} = u^t + D(\hat{\alpha}^{(i)})(v^{(i)} - v^t) \qquad (3.2)$$

where $D(\hat{\alpha}^{(i)})$ is now a 2×2 matrix. $D(\hat{\alpha}^{(i)}) = \begin{bmatrix} \hat{D}_{11}^{(i)} & \hat{D}_{12}^{(i)} \\ \hat{D}_{21}^{(i)} & \hat{D}_{22}^{(i)} \end{bmatrix}$.

From (3.2) we see that in general each of the leader's inputs $u_1^{(i)}$ and $u_2^{(i)}$ depend upon both of the follower's input deviations from the desired input. The leader may now use the follower's estimated cost functional

$$\hat{J}_F^{(i)} = u'\hat{R}_F^{(i)}u + v'\hat{S}_F^{(i)}v \qquad (3.3)$$

to calculate an estimate of the follower's response $\hat{v}^{(i)}$ to the incentive input structure $u^{(i)}$ of (3.2). This is done by finding the vector $\hat{v}^{(i)}$ which minimizes $\hat{J}_F^{(i)}$. From basic calculus and the chain rule, $\hat{v}^{(i)}$ is a solution to

$$\frac{d\hat{J}_F^{(i)}}{d\hat{v}^{(i)}} = \frac{\partial\hat{J}_F^{(i)}}{\partial\hat{v}^{(i)}} + \frac{\partial\hat{J}_F^{(i)}}{\partial u^{(i)}}\frac{\partial u^{(i)}}{\partial\hat{v}^{(i)}} = 0. \qquad (3.4)$$

We may solve (3.4) using (3.2) and (3.3) to obtain

$$\hat{v}^{(i)} = \left[[v^{t'}D(\hat{\alpha}^{(i)})'\hat{R}_F^{(i)'}D(\hat{\alpha}^{(i)}) - u^{t'}\hat{R}_F^{(i)'}D(\hat{\alpha}^{(i)})][\hat{S}_F^{(i)'} + D(\hat{\alpha}^{(i)})'\hat{R}_F^{(i)'}D(\hat{\alpha}^{(i)})]^{-1} \right]'.$$

(3.5)

The leader wants to choose $D(\hat{\alpha}^{(i)})$ such that $\hat{v}^{(i)} = v^t$. With this equality (3.5) reduces to

$$-u^{t'}\hat{R}_F^{(i)'}D(\hat{\alpha}^{(i)}) = v^{t'}\hat{S}_F^{(i)'}.$$

(3.6)

Expanding the matrices in (3.6) we obtain

$$-u_1^t(\hat{D}_{11}^{(i)} + \hat{R}_{F12}^{(i)}\hat{D}_{21}^{(i)}) - u_2^t(\hat{R}_{F12}^{(i)}\hat{D}_{11}^{(i)} + \hat{R}_{F22}^{(i)}\hat{D}_{21}^{(i)}) = v_1^t\hat{S}_{F11}^{(i)} + v_2^t\hat{S}_{F12}^{(i)}$$

(3.7a)

$$-u_1^t(\hat{D}_{12}^{(i)} + \hat{R}_{F12}^{(i)}\hat{D}_{22}^{(i)}) - u_2^t(\hat{R}_{F12}^{(i)}\hat{D}_{12}^{(i)} + \hat{R}_{F22}^{(i)}\hat{D}_{22}^{(i)}) = v_1^t\hat{S}_{F12}^{(i)} + v_2^t\hat{S}_{F22}^{(i)}$$

(3.7b)

which reveals a system of two equations and four unknowns $(\hat{D}_{11}^{(i)}, \hat{D}_{12}^{(i)}, \hat{D}_{21}^{(i)}, \hat{D}_{22}^{(i)})$. Hence our solution for $D(\hat{\alpha}^{(i)})$ is certainly non-unique and has two degrees of freedom [18].

To insure a unique solution we suppose the leader chooses $D(\hat{\alpha}^{(i)})$ to be a diagonal matrix. That is,

$$D(\hat{\alpha}^{(i)}) = \begin{bmatrix} \hat{D}_{11}^{(i)} & 0 \\ 0 & \hat{D}_{22}^{(i)} \end{bmatrix}.$$

(3.8)

Then (3.7a,b) reduces to

$$(-u_1^t - u_2^t\hat{R}_{F12}^{(i)})\hat{D}_{11}^{(i)} = v_1^t\hat{S}_{F11}^{(i)} + v_2^t\hat{S}_{F12}^{(i)}$$

(3.9a)

$$(-u_1^t\hat{R}_{F12}^{(i)} - u_2^t\hat{R}_{F22}^{(i)})\hat{D}_{22}^{(i)} = v_1^t\hat{S}_{F12}^{(i)} + v_2^t\hat{S}_{F22}^{(i)}$$

(3.9b)

and we have

$$\hat{D}_{11}^{(i)} = - \frac{v_1^t S_{F11}^{(i)} + v_2^t S_{F12}^{(i)}}{u_1^t + u_2^t R_{F12}^{(i)}} \qquad (3.10a)$$

$$\hat{D}_{22}^{(i)} = - \frac{v_1^t S_{F12}^{(i)} + v_2^t S_{F22}^{(i)}}{u_1^t R_{F12}^{(i)} + u_2^t R_{F22}^{(i)}} \qquad (3.10b)$$

as the elements of the incentive matrix $D(\hat{\alpha}^{(i)})$. The optimal diagonal incentive matrix elements D_{11}^* and D_{22}^* can be found from (3.10a,b) by using the actual values of R_{F12}, R_{F22}, S_{F11}, S_{F12}, and S_{F22}.

3.4. Parameter Estimations and Updating Techniques

The leader begins his estimation of the unknown parameters in α by making an initial guess $\hat{\alpha}^{(0)}$. Since we know that R_F and S_F are positive definite matrices it is advisable to select this initial guess $\hat{\alpha}^{(0)}$ such that the resulting matrices, $\hat{R}_F^{(0)}$ and $\hat{S}_F^{(0)}$, are both positive definite. The leader may use these parameter estimates to calculate his input structure $u^{(0)}$ via (3.2), (3.8), and (3.10a,b) and present it to the follower. The follower will optimize his cost functional (3.1b) with respect to this leader's input structure and present his input $v^{(0)}$. The subsequent value of $u^{(0)}$ may be obtained from (3.2) and each player may then compute his cost for stage zero. At this point the leader wishes to have a method for updating his parameter estimates before beginning stage one of the game. We now consider the formulation of a general updating procedure which can be used following any stage i.

The follower's true input $v^{(i)}$ at any stage i can be found by optimizing J_F in (3.1b) with respect to $v^{(i)}$ and $u^{(i)}$ as given in (3.2). By the chain rule $v^{(i)}$ is a solution to

$$\frac{dJ_F}{dv^{(i)}} = \frac{\partial J_F}{\partial v^{(i)}} + \frac{\partial J_F}{\partial u^{(i)}} \frac{\partial u^{(i)}}{\partial v^{(i)}} = 0 \tag{3.11}$$

which leads to

$$v^{(i)'} = (-u^{t'} + v^{t'} D(\hat{\alpha}^{(i)})') R_F' D(\hat{\alpha}^{(i)}) (S_F' + D(\hat{\alpha}^{(i)})' R_F' D(\hat{\alpha}^{(i)}))^{-1}. \tag{3.12}$$

This value of $v^{(i)}$ will produce a corresponding leader's input $u^{(i)}$ according to (3.2). Once these values are obtained, the leader may calculate his cost J_L for step i. At this point the leader wishes to update his estimate $\hat{\alpha}^{(i)}$ by making use of the information gained during step i. Corresponding to the first order (scalar) case, this can be done by using a gradient technique based upon minimizing J_L.

Since there are five estimated parameters in $\hat{\alpha}^{(i)}$, the leader must have five separate equations for updating his estimate. In general, we write

$$\hat{\alpha}^{(i+1)} = \hat{\alpha}^{(i)} - \gamma \nabla_{\hat{\alpha}^{(i)}} J_L \tag{3.13}$$

which is, in fact, the five equations

$$\hat{R}_{F12}^{(i+1)} = \hat{R}_{F12}^{(i)} - \gamma \nabla_{\hat{R}_{F12}^{(i)}} J_L \tag{3.14a}$$

$$\hat{R}_{F22}^{(i+1)} = \hat{R}_{F22}^{(i)} - \gamma \nabla_{\hat{R}_{F22}^{(i)}} J_L \tag{3.14b}$$

$$\hat{S}_{F11}^{(i+1)} = \hat{S}_{F11}^{(i)} - \gamma \nabla_{\hat{S}_{F11}^{(i)}} J_L \tag{3.14c}$$

$$\hat{S}_{F12}^{(i+1)} = \hat{S}_{F12}^{(i)} - \gamma \nabla_{\hat{S}_{F12}^{(i)}} J_L \tag{3.14d}$$

$$\hat{S}_{F22}^{(i+1)} = \hat{S}_{F22}^{(i)} - \gamma \nabla_{\hat{S}_{F22}^{(i)}} J_L. \tag{3.14e}$$

Expanding the matrix equations (3.1a) and (3.12), the leader may obtain the following expressions for $J_L^{(i)}$ and $v^{(i)}$ in terms of known quantities, the parameter estimates $\hat{\alpha}^{(i)}$, and the unknown parameters α. (Recall that $\hat{D}_{11}^{(i)}$ and $\hat{D}_{22}^{(i)}$ were both functions of known quantities and the estimates $\hat{\alpha}^{(i)}$.)

$$J_L = \left(\begin{array}{l} \hat{D}_{11}^{(i)^2} R_{L11}(v_1^{(i)}-v_1^t)^2 + 2\hat{D}_{11}^{(i)}\hat{D}_{22}^{(i)} R_{L12}(v_1^{(i)}-v_1^t)(v_2^{(i)}-v_2^t) + \hat{D}_{22}^{(i)^2} R_{L22}(v_2^{(i)}-v_2^t)^2 \\[2mm] + S_{L11}(v_1^{(i)}-v_1^t)^2 + 2S_{L12}(v_1^{(i)}-v_1^t)(v_2^{(i)}-v_2^t) + S_{L22}(v_2^{(i)}-v_2^t)^2 \end{array} \right) \tag{3.15}$$

$$v_1^{(i)} = \frac{\left(\begin{array}{l} [\hat{D}_{11}^{(i)}(v_1^t\hat{D}_{11}^{(i)}-u_1^t) + R_{12}\hat{D}_{11}^{(i)}(v_2^t\hat{D}_{22}^{(i)}-u_2^t)](S_{22}+R_{22}\hat{D}_{22}^{(i)^2}) \\[2mm] -[R_{12}\hat{D}_{22}^{(i)}(v_1^t\hat{D}_{11}^{(i)}-u_1^t) + R_{22}\hat{D}_{22}^{(i)}(v_2^t\hat{D}_{22}^{(i)}-u_2^t)](S_{12}+\hat{D}_{11}^{(i)}\hat{D}_{22}^{(i)}R_{12}) \end{array} \right)}{(S_{11}+\hat{D}_{11}^{(i)^2})(S_{22}+R_{22}\hat{D}_{22}^{(i)^2}) - (S_{12}+\hat{D}_{11}^{(i)}\hat{D}_{22}^{(i)}R_{12})^2} \tag{3.16a}$$

$$v_2^{(i)} = \frac{\left(\begin{array}{l} [R_{12}\hat{D}_{22}^{(i)}(v_1^t\hat{D}_{11}^{(i)}-u_1^t) + R_{22}\hat{D}_{22}^{(i)}(v_2^t\hat{D}_{22}^{(i)}-u_2^t)](S_{22}+\hat{D}_{11}^{(i)^2}) \\[2mm] -[\hat{D}_{11}^{(i)}(v_1^t\hat{D}_{11}^{(i)}-u_1^t) + R_{12}\hat{D}_{11}^{(i)}(v_2^t\hat{D}_{22}^{(i)}-u_2^t)](S_{12}+\hat{D}_{11}^{(i)}\hat{D}_{22}^{(i)}R_{12}) \end{array} \right)}{(S_{11}+\hat{D}_{11}^{(i)^2})(S_{22}+R_{22}\hat{D}_{22}^{(i)^2}) - (S_{12}+\hat{D}_{11}^{(i)}\hat{D}_{22}^{(i)}R_{12})^2}. \tag{3.16b}$$

Using the chain rule and equations (3.10), (3.15), and (3.16a,b), we can find the quantities $\nabla_{\hat{R}_{F12}^{(i)}}, \nabla_{\hat{R}_{F22}^{(i)}}, \ldots, \nabla_{\hat{S}_{F22}^{(i)}}$, through a lengthy, but straightforward process. For example,

$$\nabla_{\hat{R}_{F12}^{(i)}} J_L = \left(\frac{dJ_L}{d\hat{R}_{F12}^{(i)}}\right)' = \left[\frac{\partial J_L}{\partial \hat{D}_{11}^{(i)}}\frac{\partial \hat{D}_{11}^{(i)}}{\partial \hat{R}_{F12}^{(i)}} + \frac{\partial J_L}{\partial \hat{D}_{22}^{(i)}}\frac{\partial \hat{D}_{22}^{(i)}}{\partial \hat{R}_{F12}^{(i)}} + \frac{\partial J_L}{\partial v_1^{(i)}}\frac{\partial v_1^{(i)}}{\partial \hat{R}_{F12}^{(i)}}\right.$$

$$\left. + \frac{\partial J_L}{\partial v_2^{(i)}}\frac{\partial v_2^{(i)}}{\partial \hat{R}_{F12}^{(i)}}\right]' \qquad (3.17)$$

where $\dfrac{\partial \hat{D}_{11}^{(i)}}{\partial \hat{R}_{F12}^{(i)}}$ and $\dfrac{\partial \hat{D}_{22}^{(i)}}{\partial \hat{R}_{F12}^{(i)}}$ are found by differentiating (3.10a,b), and $\dfrac{\partial J_L}{\partial \hat{D}_{11}^{(i)}}$,

$\dfrac{\partial J_L}{\partial \hat{D}_{22}^{(i)}}$, $\dfrac{\partial J_L}{\partial v_1^{(i)}}$, and $\dfrac{\partial J_L}{\partial v_2^{(i)}}$ are found by differentiating (3.15). The

equations for $v_1^{(i)}$ and $v_2^{(i)}$ in (3.16a,b) contain the unknown parameters in

α so we must obtain $\dfrac{\partial v_1^{(i)}}{\partial \hat{R}_{F12}^{(i)}}$ and $\dfrac{\partial v_2^{(i)}}{\partial \hat{R}_{F12}^{(i)}}$ by differentiating (3.16a,b) and treating

these five unknown parameters as constants. However, our resulting expres-

sions for $\dfrac{\partial v_1^{(i)}}{\partial \hat{R}_{F12}^{(i)}}$ and $\dfrac{\partial v_2^{(i)}}{\partial \hat{R}_{F12}^{(i)}}$ will still contain these unknown parameters so

we need to replace them with our corresponding parameter estimates. After

making this substitution, all of the quantities on the right hand side of

(3.17) will be known and we can calculate $\nabla_{\hat{R}_{F12}^{(i)}} J_L$.

Calculations of the gradient of J_L with respect to each of the

other four estimated parameters are similar. Once these values are computed,

the leader may use (3.14a-e) to update his parameter estimates and obtain

$\hat{\alpha}^{(i+1)}$ which are used to begin the next stage i+1 in the game.

3.5. General Aspects of the Higher Order Problem

In this section we will attempt to address a few general questions pertaining to the higher order problem we have just studied. We will assess the performance of our algorithm in solving actual numerical problems, and discuss some implications stemming from the incentive matrix flexibility.

3.5.1. Implications of the incentive matrix flexibility

From equations (3.7a,b) we can see that there are only two equations governing the four elements of the 2×2 incentive matrix. Consequently, the optimal incentive matrix $D(\hat{\alpha}^{(i)})$ is non-unique and has two degrees of freedom. A closer inspection of (3.7a,b) reveals that $\hat{D}_{11}^{(i)}$ and $\hat{D}_{21}^{(i)}$ are isolated in one equation while $\hat{D}_{12}^{(i)}$ and $\hat{D}_{22}^{(i)}$ are isolated in the other. This decoupling allows us to solve for the columns of the incentive matrix independently of each other [18].

If we allow the incentive matrix to be a general 2×2 matrix, then each of the leader's inputs u_1 and u_2 will depend on both of the follower's inputs v_1 and v_2. However, if the incentive matrix is restricted to a diagonal structure, as it was throughout this chapter, then there is a slight partitioning of the control space. In this case u_1 depends only upon v_1 and u_2 depends only upon v_2.

Although we selected a diagonal incentive matrix for simplicity and to insure a unique solution, there are certainly other meaningful ways to exercise the two degrees of freedom. One of the most useful methods would be to select the optimal incentive matrix in a manner such that the optimal incentive control is robust to parameter estimation errors [19].

Specifically, the leader would like to minimize the sensitivity of his cost functional J_L to variations in the parameter estimates $\hat{\alpha}^{(i)}$. This method has been studied for the deterministic problem [11], and its ideas should be extendable to the problem at hand.

To minimize the estimation procedure, we also investigated the possibility of limiting our estimations strictly to the elements of the incentive matrix D. This approach was appealing because in the case of a diagonal incentive matrix it reduced the number of elements of $\hat{\alpha}$ from five ($\hat{R}_{F12}$, $\hat{R}_{F22}$, $\hat{S}_{F11}$, $\hat{S}_{F12}$, and $\hat{S}_{F22}$) to two ($\hat{D}_{11}$ and $\hat{D}_{22}$). Even with the most general form of the incentive matrix the number of elements of $\hat{\alpha}$ is reduced from five to four. Although the concept seemed promising, we were unable to find a suitable method for directly updating the estimates of $\hat{D}_{11}$ and $\hat{D}_{22}$ due to a lack of information about the unknown and unestimated system parameters. Therefore, we adopted our present approach of attempting to estimate all of the unknown system parameters, and then constructing the optimal incentive matrix at each stage as a function of these estimates.

3.5.2. Performance analysis of the algorithm

Unlike the scalar case, we have thus far been unable to analytically prove that the iterative method described in this chapter will indeed produce a sequence of incentive control structures which converge to the optimal incentive control. Nevertheless, extensive simulation studies utilizing this iterative method have produced some good convergence results.

Since our parameter estimates are updated by using a gradient method on J_L, we expect the value of J_L to decrease at each successive stage of the game. We also expect J_L to converge to its optimal value of

zero unless it settles at a local minimum. Unfortunately, we must remark that we cannot be totally assured of a monotonically decreasing J_L, because strictly speaking, our updating technique is not a true gradient technique. Recall that our calculation of expressions such as $\dfrac{\partial v_1^{(i)}}{\partial \hat{R}_{F12}^{(i)}}$ and $\dfrac{\partial v_2^{(i)}}{\partial \hat{R}_{F12}^{(i)}}$ used in the updating technique involved some of the unknown system parameters α. This prevented us from calculating expressions which were crucial to the updating process. To circumvent this problem, we replaced the unknown parameters α with their most recent estimates $\hat{\alpha}^{(i)}$, enabling us to calculate the necessary expressions and continue our iterative process. This substitution is of small consequence provided the values of $\hat{D}_{11}^{(i)}$ and $\hat{D}_{22}^{(i)}$ computed from the estimates $\hat{\alpha}^{(i)}$ are close to their optimal values. It does create greater difficulties in cases where the computed values of $\hat{D}_{11}^{(i)}$ and $\hat{D}_{22}^{(i)}$ are less accurate estimates of their optimal values.

Our simulation studies appear to confirm the "near-gradient" nature of our updating scheme. For a relatively good initial guess $\hat{\alpha}^{(0)}$, i.e., a guess such that $\hat{D}_{11}^{(0)}$ and $\hat{D}_{22}^{(0)}$ are "reasonably close" to their optimal values D_{11}^{*} and D_{22}^{*}, our scheme does indeed behave similar to a gradient technique and the resulting leader's cost J_L decreases monotonically to zero. However, for a less fortunate initial guess $\hat{\alpha}^{(0)}$, our scheme deviates from the behavior of a gradient technique and does not produce a monotonically decreasing J_L. In some of these cases our scheme actually exhibits a behavior characteristic of dual control [12] by sacrificing the minimization of J_L over an initial iterative period in order to gain better estimates of the unknown parameters. Once the scheme achieves parameter estimates which provide sufficiently accurate estimates of $\hat{D}_{11}^{(i)}$ and $\hat{D}_{22}^{(i)}$,

it then proceeds to generate control inputs which result in a monotonically decreasing J_L. An example of this phenomenon is illustrated in Figure 3.1. From this figure we can see that the leader's cost actually increases for approximately the first ten iterations before beginning its monotonic decent towards zero.

Simulation studies have also revealed that our scheme will not necessarily provide accurate estimates $\hat{\alpha}^{(i)}$ for the actual values of α. In cases where our scheme successfully generates the optimal incentive control, the leader's cost functional approaches zero and the generated values $\hat{D}_{11}^{(i)}$ and $\hat{D}_{22}^{(i)}$ approach their optimal values D_{11}^* and D_{22}^*, respectively. Consequently, the follower's actual inputs $v_1^{(i)}$ and $v_2^{(i)}$ also approach their desired values of v_1^t and v_2^t. In general though, the estimates $\hat{\alpha}^{(i)}$ do not necessarily approach α. In fact, the estimates of some elements of α may converge to values far from their actual value. From this we can see that our scheme is not actually attempting to estimate α, but merely uses the estimates $\hat{\alpha}^{(i)}$ as a vehicle through which it can compute and adjust the estimations $\hat{D}_{11}^{(i)}$ and $\hat{D}_{22}^{(i)}$. Thus, although we were unable to accurately estimate the elements of the optimal diagonal incentive matrix by themselves, as discussed in the previous section, we are able to accurately estimate these elements by expressing them as functions of estimates of the unknown system parameters.

The selection of the step size γ is crucial in applying our iterative scheme. It must be chosen small enough to avoid large overshoots in the parameter estimations, but large enough to provide a sufficient adjustment in the incentive matrix. This will insure an adequate rate of convergence. In the event that the scheme becomes bogged down at an
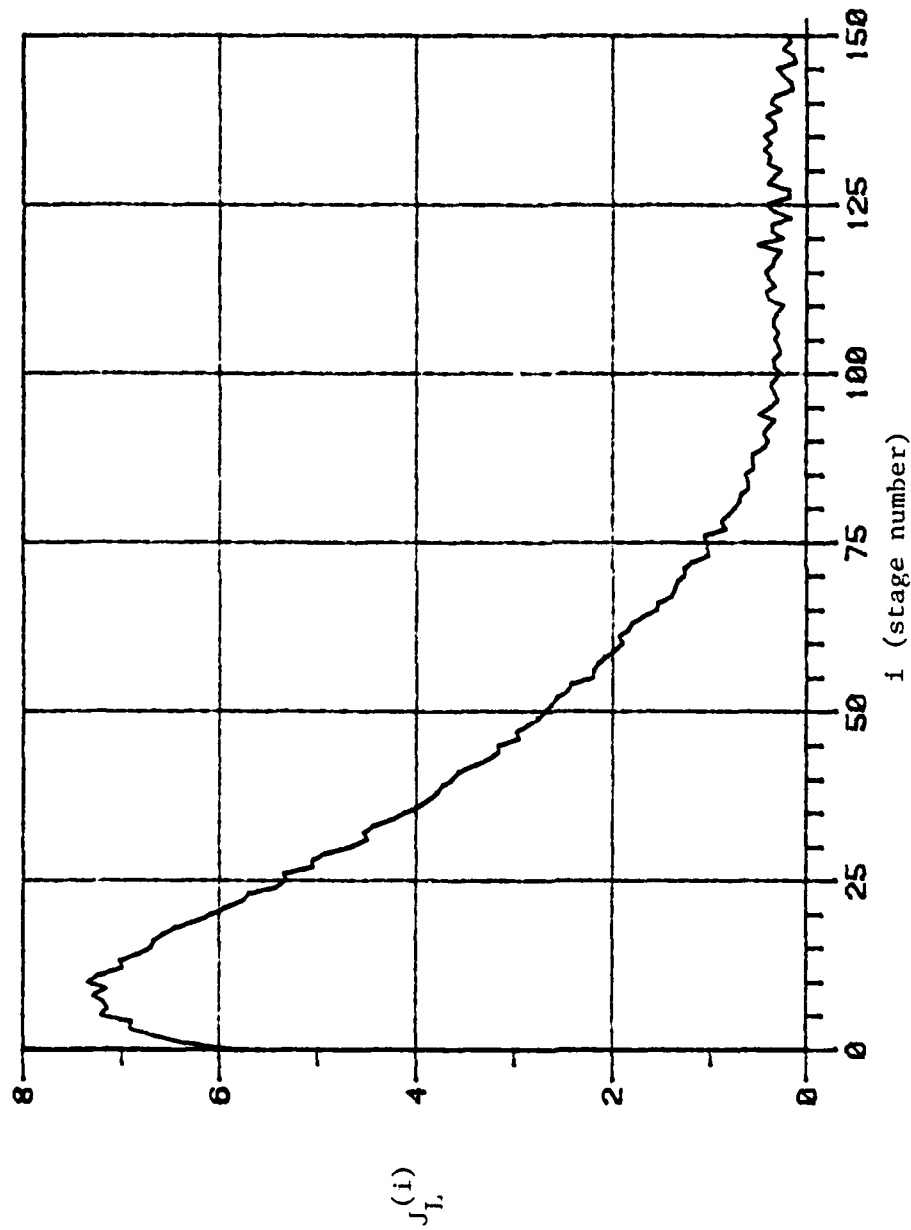
35

Figure 3.1. Time response of the leader's cost $J_L^{(i)}$ demonstrating the possible "dual effect" behavior of our algorithm.

undesired local minimum, the authors propose restarting the algorithm with

an entirely new initial guess $\hat{\alpha}^{(0)}$. No recommendations are made regarding

restarting the algorithm with an initial guess $\hat{\alpha}^{(0)}$ modified from the

previously unsuccessful initial guess.

3.5.3. Generalization of the algorithm to higher dimensional problems

In this chapter we have developed an algorithm which iteratively

computes incentive input controls for the leader in a static, second

order Stackelberg game with unknown cost functionals. This algorithm was

developed without using any inherent properties of the second order problem,

leading us to believe that it can be extended to a general n-th order

problem. In a general n-th order problem, equations (3.7a,b) will represent

a system of n equations and n^2 unknowns. This results in n(n-1) degrees of

freedom in selecting the optimal incentive matrix. We can once again be

assured of a unique solution by considering only diagonal incentive

matrices. Furthermore, we believe that this solution can be computed by

applying a generalized algorithm similar to the one described in this

chapter.

CHAPTER 4

SIMULATION EXAMPLES

In the previous two chapters we have discussed methods for deriving optimal incentive controls for two-player Stackelberg games with unknown cost functionals. We now demonstrate the results of these methods via a few numerical examples. The first example is a realistic economic problem which uses the theory of Chapter 2 and implements the error function method for updating parameter estimations. The second example demonstrates the algorithm described in Chapter 3 applied to a second order problem.

4.1. A Scalar Economic Example

Consider the following economic problem [20,21] illustrated in Figure 4.1.

A monopoly M operates in a market with a demand curve specified by $p = A_1 - A_2 q$ and with a flat marginal cost curve $MC = C$ dollars/unit. The government, which does not know the value of the parameters A_1 or A_2, wishes to regulate this monopoly in such a way that its production output will be equal to q^*, the same quantity which would be produced in a purely competitive market. We do assume, through market surveys or estimates, that the government does have knowledge of the current operating point (q_m, p_m). The government regulation may be either a tax or a subsidy, and can be applied in either a lump sum or per quantity method.

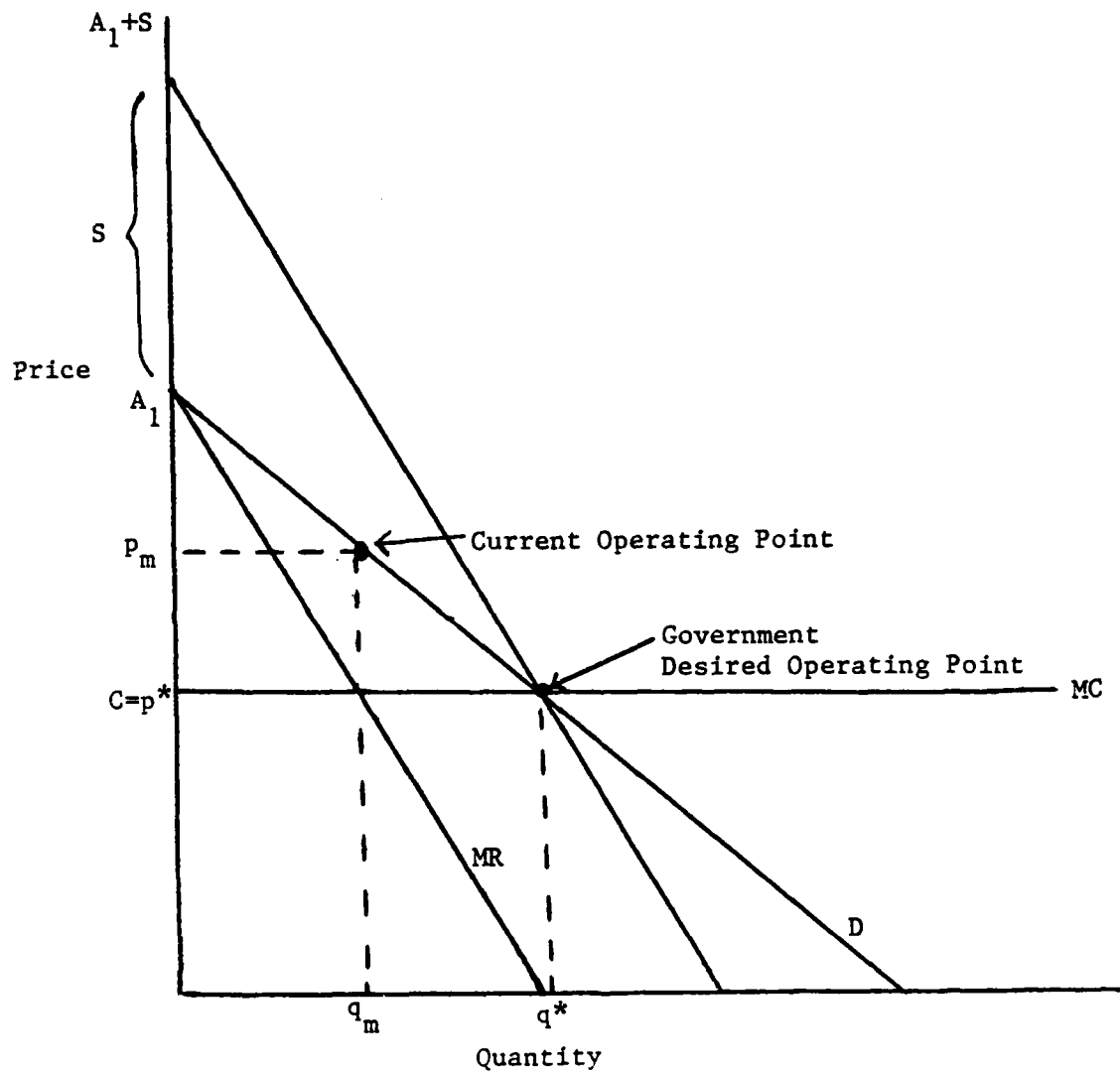Let p = price and q = quantity produced. We have a demand curve specified by

Figure 4.1. Government regulation of a monopoly.

$$p = A_1 - A_2 q. \tag{4.1}$$

By economic theory [21] we know that the marginal revenue (MR) curve has the same intercept and twice the slope of the demand curve. Thus,

$$MR = A_1 - 2A_2 q. \tag{4.2}$$

Since we know that (q_m, p_m) must be a point on the demand curve, we can express A_1 as a function of A_2 and thus eliminate one of the unknown parameters. From (4.1) we have

$$A_1 = p_m + A_2 q_m. \tag{4.3}$$

We know that the monopolist will produce at a level where $MR = MC$. Let us assume that the government decides to provide a subsidy of S dollars per unit produced. If A_2 were known, then

$$MR = MC \implies A_1 + S - 2A_2 q^* = A_1 - A_2 q^* \implies S = A_2 q^* \tag{4.4}$$

would be the optimum subsidy. However, A_2 is unknown so the government will attempt to estimate it with $\hat{A}_2^{(i)}$. Consider the incentive structured subsidy

$$S = \hat{A}_2^{(i)} q^* + D(\hat{A}_2^{(i)})(q - q^*). \tag{4.5}$$

Given this subsidy, the monopolist will still produce at the level of ouptut which maximizes his profit. To estimate this level of production $\hat{q}^{(i)}$, the government solves (using $\hat{A}_2^{(i)}$ as an estimate for A_2)

$$\frac{d\hat{J}_F^{(i)}}{\partial \hat{q}^{(i)}} = 0 \tag{4.6}$$

where

$$\hat{J}_F^{(i)} = \hat{q}^{(i)}(\text{PRICE-COST}) = \hat{q}^{(i)}(\hat{A}_2 q_m + p_m - \hat{A}_2 \hat{q}^{(i)}) \tag{4.7}$$

is the monopolist's total profit. Solving (4.6) by using (4.7) we obtain

$$\hat{q}^{(i)} = \frac{C + D(\hat{A}_2^{(i)})q^* - \hat{A}_2^{(i)}q^* - p_m - \hat{A}_2^{(i)}q_m}{2(D(\hat{A}_2^{(i)}) - \hat{A}_2^{(i)})} \tag{4.8}$$

as an estimate for the monopolist's production level. Setting $\hat{q}^{(i)}$ in (4.8) equal to the desired output q^*, the government calculates its optimal incentive constant to be

$$D(\hat{A}_2^{(i)}) = \frac{\hat{A}_2^{(i)}(q^*-q_m) + C - p_m}{q^*} . \tag{4.9}$$

However, since our estimate $\hat{A}_2$ of A_2 is probably incorrect, the monopolist's true production level will probably differ from q^*. Solving (4.6) and (4.7) with the actual value of A_2, we find that the monopolist actually produces

$$q^{(i)} = \frac{(C - p_m - \hat{A}_2 q_m)q^*}{(C - p_m + \hat{A}_2(q^*-q_m) - A_2 q^*)} \text{ units.} \tag{4.10}$$

Now consider the positive definite error function

$$E_i = \frac{1}{2}(q^{(i)}-q^*)^2. \tag{4.11}$$

The government may update its estimates for $\hat{A}_2^{(i)}$ by using a gradient method on E_i

$$\hat{A}_2^{(i+1)} = \hat{A}_2^{(i)} - \gamma \nabla_{\hat{A}_2^{(i)}} E_i = \hat{A}_2^{(i)} - \gamma(q^{(i)}-q^*)\nabla_{\hat{A}_2^{(i)}} q^{(i)}. \tag{4.12}$$

Differentiating (4.10) and using $\hat{A}_2$ as an estimate for A_2, we have

$$\nabla_{\hat{A}_2^{(i)}} q^{(i)} = \frac{-q^{*2}}{(C - p_m - \hat{A}_2 q_m)} \tag{4.13}$$

which is substituted into (4.12) to generate the final form of the updating equation

$$\hat{A}_2^{(i+1)} = \hat{A}_2^{(i)} + \gamma(q^{(i)}-q^*) \left[\frac{q^{*2}}{(C - p_m - \hat{A}_2 q_m)} \right]. \tag{4.14}$$

In this example let our variables take on the following numerical values

$A_1 = \$450$

$A_2 = \$1.50/\text{unit}$

$q_m = 130 \text{ units}$

$p_m = \$255/\text{unit}$

$p^* = \text{cost } C = \$60/\text{unit}$

$q^* = 260 \text{ units}$

$\gamma = .00001$

$\text{sigma} = 3.0.$

Then the updating equation (4.14) takes the form

$$\hat{A}_2^{(i+1)} = \hat{A}_2^{(i)} + \gamma(q^{(i)}-260) \left(\frac{260^2}{-195-130\hat{A}_2^{(i)}} \right). \tag{4.15}$$

Assume that the government makes an initial guess of $\hat{A}_2^{(0)} = .65$. By simulation with a small amount (35dB S/N ratio) of noise, we obtain the results illustrated in Figures 4.2-4.4. Figure 4.2 illustrates the estimate $\hat{A}_2^{(i)}$ at each stage i. In this case $\hat{A}_2^{(i)}$ converges to the actual value of A_2 rather quickly (in approximately twenty iterations). Using these estimates $\hat{A}_2^{(i)}$ to implement the incentive structured subsidy computed from (4.5) and (4.9), the resulting quantity produced and market price at each stage i are displayed in Figures 4.3 and 4.4, respectively.
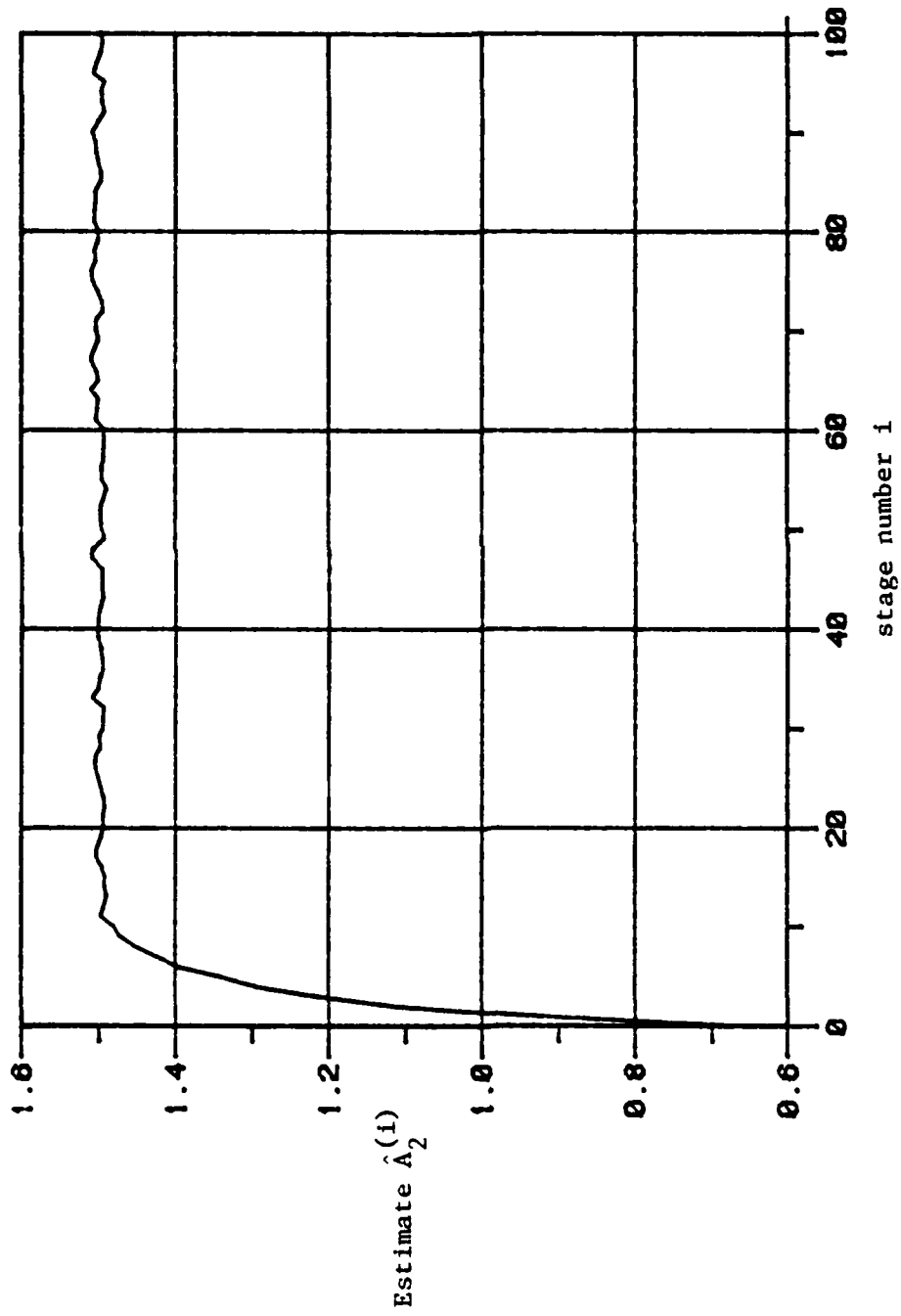
Figure 4.2. Time response of the parameter estimate $\hat{A}_2^{(i)}$.
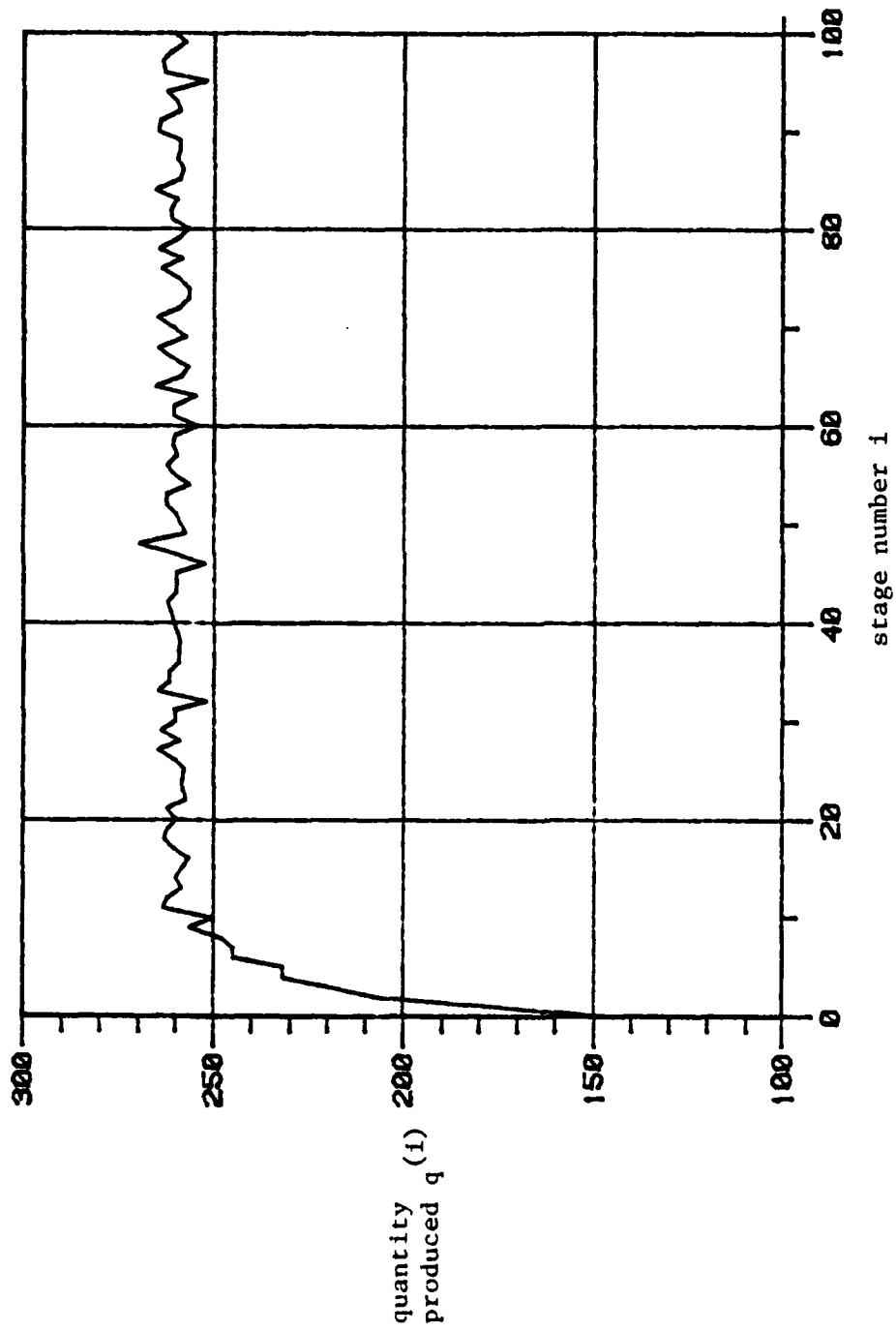
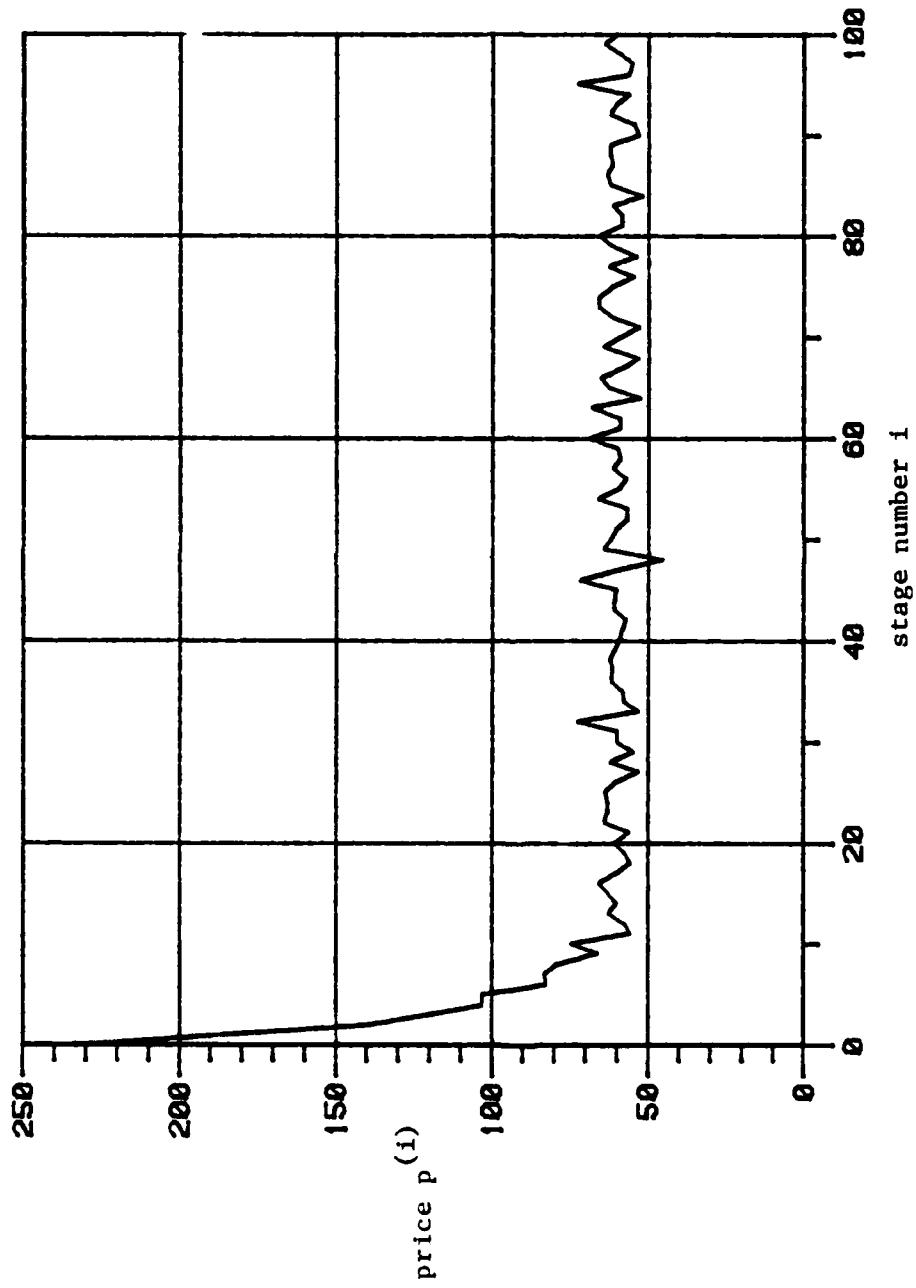Figure 4.3. Time response of the quantity of goods produced $q^{(1)}$.

Figure 4.4. Time response of the market price $p^{(i)}$.

4.2. A Second Order Example

Consider the following second order numerical example

$$J_L = (u-u^t)'R_L(u-u^t) + (v-v^t)'S_L(v-v^t) \qquad (4.16a)$$

$$J_F = u'R_F u + v'S_F v \qquad (4.16b)$$

where $\qquad u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ is the leader's 2×1 input vector

and $\qquad v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ is the follower's 2×1 input vector.

Suppose we have the following values

$$u^t = \begin{bmatrix} 2.8 \\ 3.4 \end{bmatrix} \qquad\qquad v^t = \begin{bmatrix} 2.15 \\ 4.6 \end{bmatrix} \qquad (4.17)$$

$$R_L = \begin{bmatrix} 1.0 & .63 \\ .63 & 1.8 \end{bmatrix} \qquad S_L = \begin{bmatrix} .86 & 1.1 \\ 1.1 & 2.5 \end{bmatrix} \qquad (4.18)$$

$$R_F = \begin{bmatrix} 1.0 & .75 \\ .75 & 1.2 \end{bmatrix} \qquad S_F = \begin{bmatrix} .75 & .65 \\ .65 & 1.6 \end{bmatrix} \qquad (4.19)$$

and the leader, who does not know the contents of the matrices R_F and S_F, wishes to apply an optimal incentive input

$$u^{(i)} = u^t + D(\hat{\alpha}^{(i)})(v^{(i)}-v^t) \qquad (4.20)$$

where $D(\hat{\alpha}^{(i)})$ is a diagonal matrix.

From equations (3.10a,b) and (4.19) we find the optimal diagonal incentive matrix elements to be

$$D_{11}^* = -.8603 \qquad\qquad D_{22}^* = -1.417. \qquad (4.21)$$

Thus, $D^* = \begin{bmatrix} -.8603 & 0 \\ 0 & -1.417 \end{bmatrix}$ is the optimal diagonal incentive matrix. Using

the algorithm described in Chapter 3 along with the initial guesses

$$\hat{R}_F^{(0)} = \begin{bmatrix} 1.0 & 1.1 \\ 1.1 & 2.0 \end{bmatrix} \qquad \hat{S}_F^{(0)} = \begin{bmatrix} 1.5 & .8 \\ .8 & 1.0 \end{bmatrix} \qquad (4.22)$$

and a step size $\gamma = .002$, and applying at each stage i the incentive

control calculated from (3.2), (3.8), and (3.10), we obtained the simula-

tion results displayed in Figures 4.5 – 4.9. These results include a small

amount (45 dB signal/noise ratio) of noise. Figure 4.5 displays the leader's

cost $J_L^{(i)}$ incurred at each stage, while Figures 4.6-4.7 illustrate the

values of $\hat{D}_{11}^{(i)}$ and $\hat{D}_{22}^{(i)}$ calculated at each stage. We can see that J_L

approaches its optimal value of zero and $\hat{D}_{11}^{(i)}$ and $\hat{D}_{22}^{(i)}$ each approach their

optimal values. The resulting components $v_1^{(i)}$ and $v_2^{(i)}$ of the follower's

input are displayed in Figures 4.8 and 4.9, respectively. These plots also

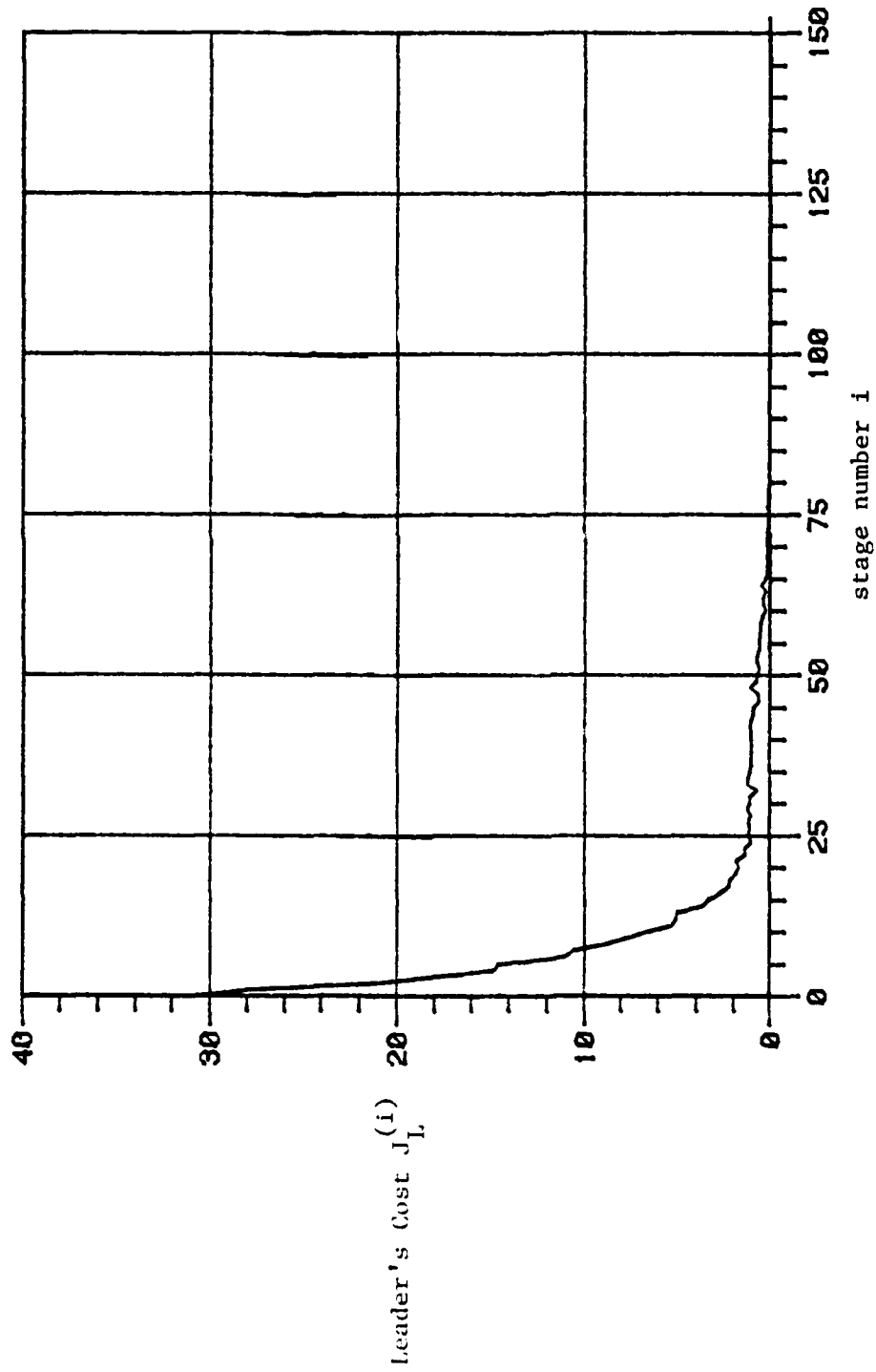approach their desired values of v_1^t and v_2^t.

47

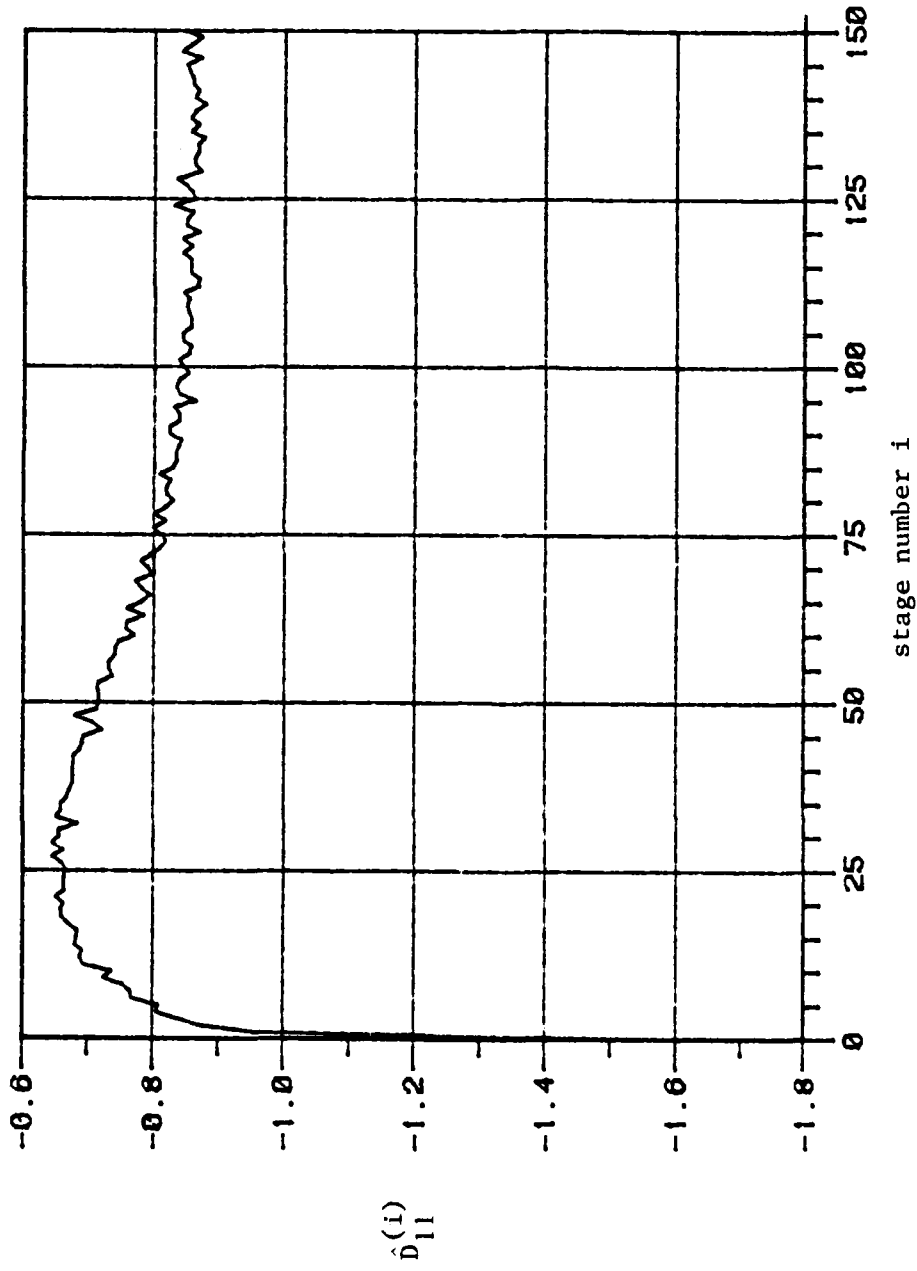Figure 4.5. Time response of the leader's cost $J_L^{(i)}$.

Figure 4.6. Time response of the incentive matrix element $\hat{D}_{11}^{(i)}$.
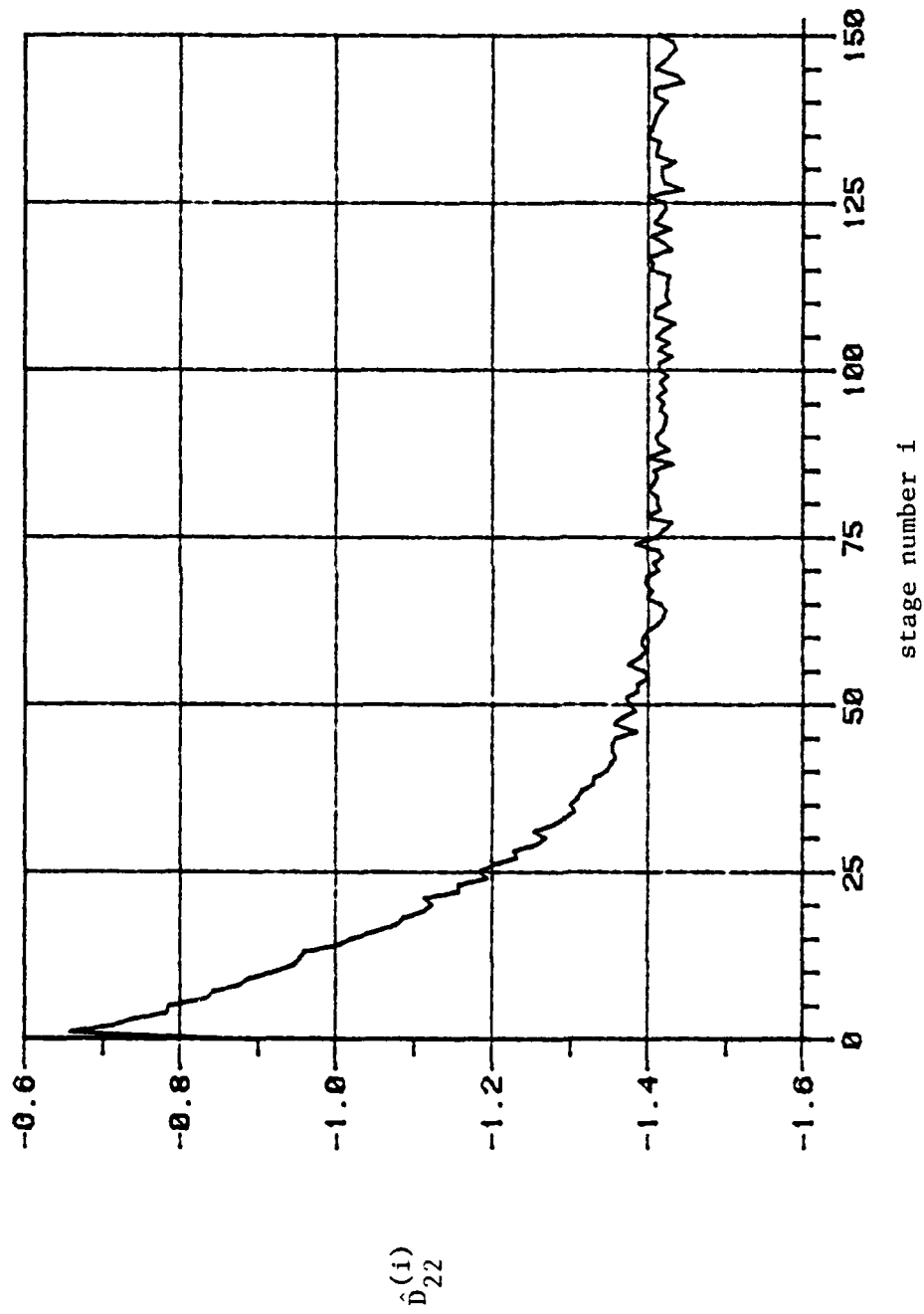
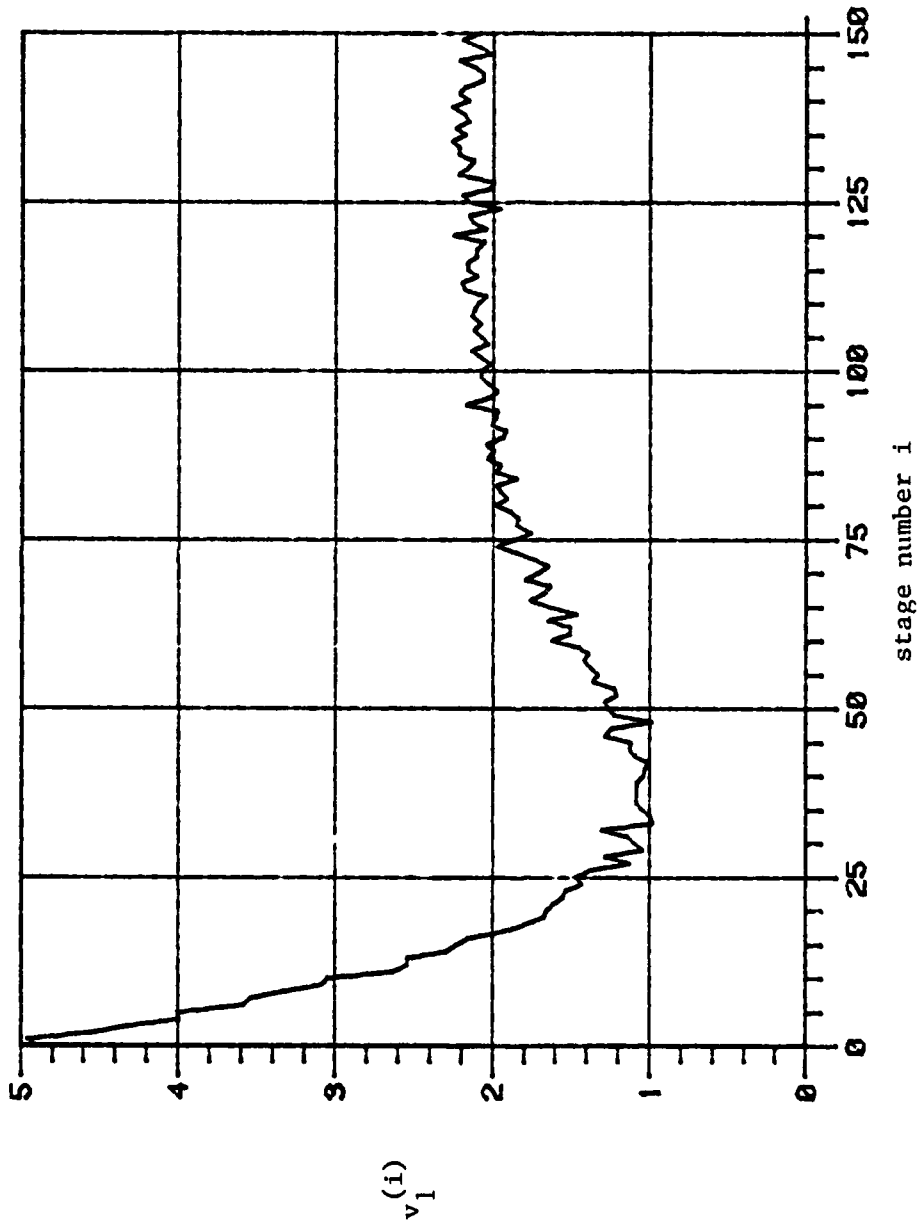Figure 4.7. Time response of the incentive matrix element $\hat{D}_{22}^{(i)}$.

50

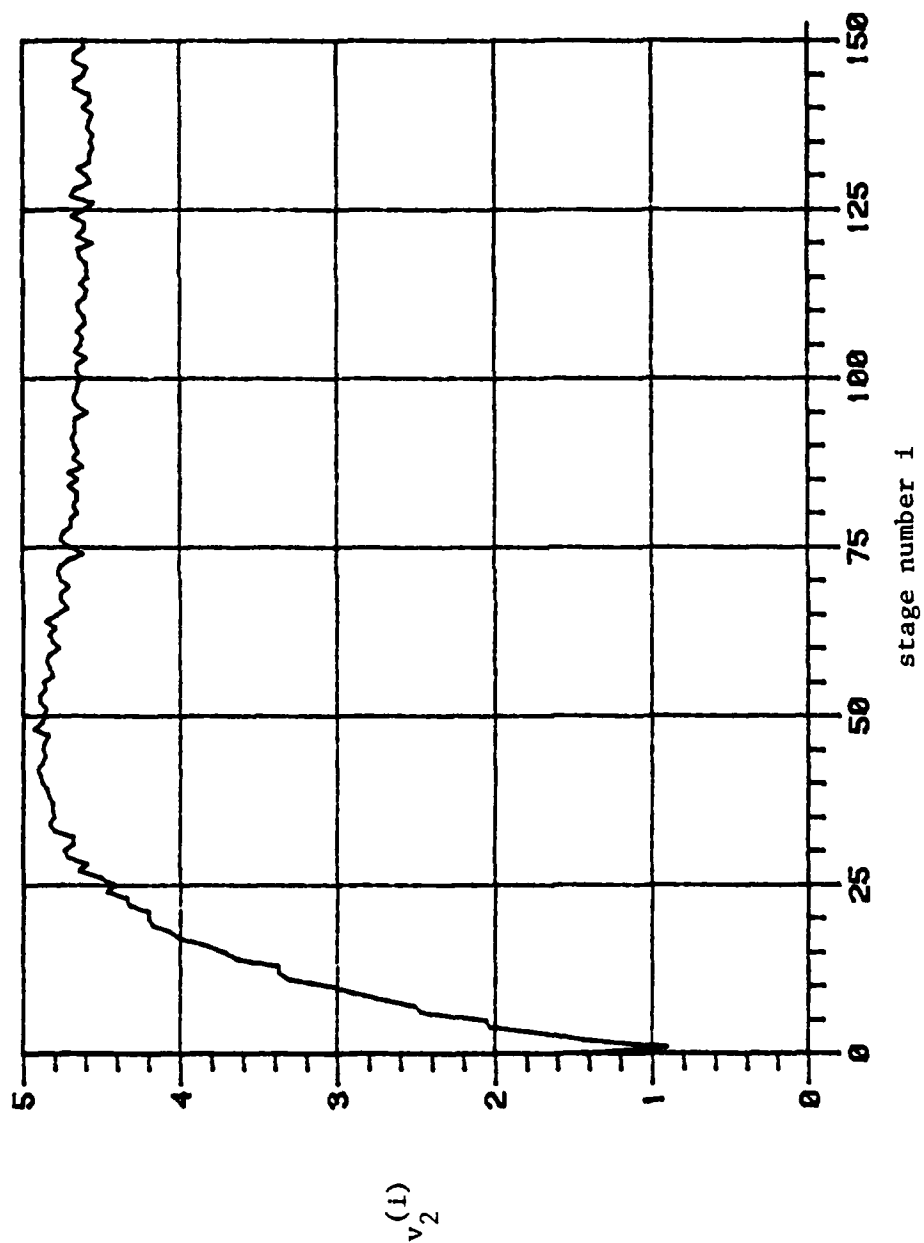Figure 4.8. Time response of the follower's input $v_1^{(i)}$.

Figure 4.9. Time response of the follower's input $v_2^{(1)}$.

CHAPTER 5

SUMMARY AND CONCLUDING REMARKS

In this thesis ~~we have~~ used the certainty equivalence approach and the theory of self-tuning regulators to derive an iterative method which generates an optimal incentive control for the leader in a static, two-player Stackelberg game with unknown cost functionals. ~~Our~~ The method uses all available degrees of freedom to restrict the incentive matrix to a diagonal structure. This restriction assures the leader of a unique optimal incentive control. Convergence to the optimal incentive control has been proven for the scalar problem and simulation studies have shown good convergence results for the second order problem. ~~We fully believe~~ that this method is extendable in its present form to a general n-th order problem.

In Chapter 4 we applied our iterative method to a scalar economic example involving government regulation of a monopoly. A simulation study of the problem revealed that the desired regulation was indeed achieved. We also demonstrated the effectiveness of ~~our~~ method on a general second order numerical problem.

Future research regarding application of optimal incentive controls to Stackelberg games with unknown cost functionals may now focus on two general areas. Starting with the iterative method detailed in this thesis, one may abandon the diagonal incentive matrix structure and attempt to use the resulting degrees of freedom to satisfy other useful criteria. An example of this is given by the minimum sensitivity design approach mentioned earlier. It is ~~also quite~~ desirable to ~~try and~~ extend the existing methods to dynamical systems and to problems involving more than two players.

REFERENCES

1. Athans, M. and P. L. Falb, Optimal Control, McGraw-Hill, New York, 1966.

2. Bertsekas, D. P., Dynamic Programming and Stochastic Control, Academic Press, New York, 1976.

3. Brogan, W. L., Modern Control Theory, Quantum Publishers, New York, 1974.

4. Bryson, A. E. and Y. C. Ho, Applied Optimal Control, Blaisdell Publishing Co., Waltham, MA, 1969.

5. Kwakernaak, H. and R. Sivan, Linear Optimal Control Systems, Wiley-Interscience, New York, 1972.

6. Khalil, H. and P. V. Kokotovic, "Control Strategies for Decision Makers Using Different Models of the Same System," IEEE Trans. on Automatic Control, vol. AC-23, April 1978, pp. 289-297.

7. Basar, T. and J. B. Cruz, Jr., "Concepts and Methods in Multi-Person Coordination and Control," Coordinated Science Lab. Report R-920, University of Illinois, Urbana, October 1981.

8. Cruz, J. B., Jr., "Leader-Follower Strategies for Multilevel Systems," IEEE Trans. on Automatic Control, vol. AC-23, April 1978, pp. 244-255.

9. Zheng, Y. P., T. Basar, and J. B. Cruz, Jr., "Stackelberg Strategies and Incentives in Multi-Person Deterministic Decision Problems," IEEE Trans. on Systems, Man, and Cybernetics, January 1984, to appear.

10. Ho, Y. C., P. B. Luh, and G. J. Olsder, "A Control Theoretic View on Incentives," Automatica, vol. 18, 1982, pp. 167-180.

11. Cansever, D. H. and T. Basar, "A Minimum Sensitivity Approach to Incentive Design Problems," Proc. of 21st IEEE Conf. on Decision and Control, Orlando, FL, December 1982, pp. 158-163.

12. Wittenmark, B., "Stochastic Adaptive Control Methods: A Survey," International Journal of Control, vol. 21, No. 5, 19 , pp. 705-730.

13. Astrom, K. J., U. Borrison, L. Ljung, and B. Wittenmark, "Theory and Applications of Self-Tuning Regulators," Automatica, vol. 13, 1977, pp. 457-476.

14. Chan, Y. M. and J. B. Cruz, Jr., "Self-Tuning Leader-Follower Games," Automatica, vol. 19, 1983, pp. 237-245.

15. Chan, Y. M., "Self-Tuning Methods for Multiple-Controller Systems," Coordinated Science Laboratory Report R-915, University of Illinois, Urbana, August 1981.

16. Ljung, L. and E. Trulsson, "Adaptive Control Based on Explicit Criterion Minimization," Department of Electrical Engineering, Linkoping Univ., Sweden, Report LiTH-ISY-I-0401, September 1980.

17. Pachter, M., "Decentralized Linear-Quadratic Differential Games with Incentives," Preprints of the 4th IFAC/IFORS/IIASA Conf. on the Modelling and Control of National Economies, Washington, D. C., June 1983, pp. 269-275.

18. Strang, G., Linear Algebra and its Applications, Academic Press, New York, 1980.

19. Cansever, D. H., T. Basar, and J. B. Cruz, Jr., "Robustness of Incentive Policies in Team Problems with Discrepancies in Goal Perceptions," Preprints of the 4th IFAC/IFORS/IIASA Conf. on the Modelling and Control of National Economies, Washington, D. C., June 1983, pp. 275-280.

20. Salman, M. and J. B. Cruz, Jr., "An Incentive Model of Duopoly with Government Coordination," Automatica, vol. 17, 1981, pp. 821-829.

21. Henderson, J. M. and R. E. Quandt, Microeconomic Theory: A Mathematical Approach, McGraw-Hill, New York, 1971.

END

FILMED

1-86

DTIC