


Article

Dynamic Mechanism Design for Repeated Markov Games with Hidden Actions: Computational Approach

Julio B. Clempner 

Escuela Superior de Física y Matemáticas, Instituto Politécnico Nacional, School of Physics and Mathematics, National Polytechnic Institute, Edificio 9 U.P. Adolfo Lopez Mateos, Col. San Pedro Zacatenco, Mexico City 07730, Mexico; julio@clempner.name

Abstract: This paper introduces a dynamic mechanism design tailored for uncertain environments where incentive schemes are challenged by the inability to observe players' actions, known as moral hazard. In these scenarios, the system operates as a Markov game where outcomes depend on both the state of payouts and players' actions. Moral hazard and adverse selection further complicate decision-making. The proposed mechanism aims to incentivize players to truthfully reveal their states while maximizing their expected payoffs. This is achieved through players' best-reply strategies, ensuring truthful state revelation despite moral hazard. The revelation principle, a core concept in mechanism design, is applied to models with both moral hazard and adverse selection, facilitating optimal reward structure identification. The research holds significant practical implications, addressing the challenge of designing reward structures for multiplayer Markov games with hidden actions. By utilizing dynamic mechanism design, researchers and practitioners can optimize incentive schemes in complex, uncertain environments affected by moral hazard. To demonstrate the approach, the paper includes a numerical example of solving an oligopoly problem. Oligopolies, with a few dominant market players, exhibit complex dynamics where individual actions impact market outcomes significantly. Using the dynamic mechanism design framework, the paper shows how to construct optimal reward structures that align players' incentives with desirable market outcomes, mitigating moral hazard and adverse selection effects. This framework is crucial for optimizing incentive schemes in multiplayer Markov games, providing a robust approach to handling the intricacies of moral hazard and adverse selection. By leveraging this design, the research contributes to the literature by offering a method to construct effective reward structures even in complex and uncertain environments. The numerical example of oligopolies illustrates the practical application and effectiveness of this dynamic mechanism design.

Keywords: optimal dynamic mechanism design; hidden actions; Markov games with private information; Bayesian equilibrium; regularization



Citation: Clempner, J.B. Dynamic Mechanism Design for Repeated Markov Games with Hidden Actions: Computational Approach. *Math. Comput. Appl.* **2024**, *29*, 46. <https://doi.org/10.3390/mca29030046>

Academic Editor: Oliver Schütze

Received: 15 April 2024

Revised: 30 May 2024

Accepted: 6 June 2024

Published: 10 June 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Brief Review

In recent decades, theoretical literature has increasingly emphasized the mathematical modeling and analysis of incentive schemes, also referred to as principal-player relationships [1], particularly in uncertain environments. In such contexts, parties involved face constraints arising from four main cases: (a) risk-sharing with symmetric information, (b) hidden type with asymmetric information, (c) hidden type with partially observable states involving asymmetric information, and (d) hidden action due to the inability to observe each player's actions. The second and fourth constraints, known in economic theory as adverse selection and moral hazard, respectively, are of particular interest. The central challenge posed by these constraints revolves around designing mechanisms that effectively address adverse selection and moral hazard while maximizing rewards [2–4]. Adverse selection refers to situations where one party possesses information not available to others, potentially leading to suboptimal outcomes. Moral hazard, on the other hand,

arises when one party's actions are not observable, resulting in risks that are not fully accounted for in the incentive structure. Addressing these issues requires a sophisticated mechanism design that aligns the interests of all parties involved and mitigates the adverse effects of asymmetric information. By focusing on the design of mechanisms that incentivize truthful disclosure of information and discourage opportunistic behavior, researchers aim to develop strategies that maximize rewards while ensuring efficiency and fairness in economic interactions. This involves creating incentive schemes that align individual actions with desirable outcomes, effectively balancing the risks and asymmetries in information. The objective is to establish a framework where parties are motivated to act in ways that lead to mutually beneficial results, even in the presence of hidden information and actions.

The classical approach to hidden action, as outlined by Cvitanic [5], pertains to situations where a player's actions remain unobservable to the principal. This lack of observability leads to a reduction in the principal's expected profit, as the consequences of the player's hidden actions may not align with the principal's objectives. Real-world applications are abundant where principals cannot infer players' actions due to the high cost of observation or inherent impossibility. Consider the insurance market, where the expectations placed on insurance companies are heavily influenced by the moral and cultural values of insured individuals. Moral hazard manifests when an insured individual exerts less effort in maintaining their health, leading to increased costs for the insurer. This happens because the insured party does not bear the full financial consequences of their actions, therefore incentivizing behavior that raises costs for the insurer, such as unnecessary medical services. Adverse selection, on the other hand, relates to inherent characteristics of the insured individual, such as genetic predispositions. In this scenario, individuals with higher genetic risks are more likely to seek insurance, creating a situation where the pool of insured individuals is disproportionately composed of higher-risk individuals. This leads to adverse outcomes for the insurer, as they may face higher-than-expected costs due to the concentration of high-risk individuals in their customer base. Overall, both moral hazard and adverse selection underscore the challenges faced by principals in designing effective incentive mechanisms and managing risk in environments where player actions are hidden or difficult to observe. Addressing these challenges requires innovative approaches that account for asymmetric information and align incentives to achieve desired outcomes. This involves creating mechanisms that not only incentivize desired behaviors but also manage the risks associated with unobservable actions and hidden information.

When actions are unobservable or non-contractable, the principal encounters difficulty in selecting the desired actions for the player. Instead, the principal must identify the optimal actions for the player and devise a contract that encourages desirable behavior. This process revolves around incentives, with the principal seeking to influence the player's choices by crafting an appropriate contract. This situation epitomizes moral hazard, wherein players may act contrary to the principal's interests. This underscores the significance of creating effective incentive mechanisms to mitigate such risks and ensure alignment between the player's actions and the principal's objectives.

In numerous instances documented in the literature, the principal lacks private information, while players possess private information and can take actions hidden from the principal. This scenario gives rise to both adverse selection and moral hazard. Conversely, when players lack private information, moral hazard arises with *ex ante* (risk-sharing with) symmetric uncertain information. The distinction between these scenarios is crucial for comprehending the dynamics of incentive structures. Furthermore, the state of the system is categorized as *exogenous* if players' actions do not directly influence the payoff. Conversely, if players' actions affect the payoff, the state is deemed *endogenous*. These categorizations, elucidated by Garrett et al. [6], Board et al. [7], and Halac et al. [8], offer a framework for analyzing the interplay between information asymmetry, strategic behavior, and the design of incentive mechanisms. Understanding these dynamics is pivotal for devising effective

strategies to mitigate adverse effects such as moral hazard and adverse selection in various economic and organizational contexts.

Fudenberg et al. [9] pioneered a dynamic principal-player model centered on a stochastic process with hidden actions, advocating for the breakdown of optimal long-term contracts into easily computable short-term contracts. Plambeck and Zenios [10] extended this study by integrating the principal-player model with the physical structure of a Markov decision process, providing a dynamic model that captures the intricate interplay between hidden actions and stochastic processes. Cole and Kocherlakota [11] further advanced this field by proposing a solution for dynamic Markov games where each player's actions remain unobservable to others, and these actions can influence an unobservable state variable. This enriched the understanding of strategic interactions in dynamic settings.

In the context of Markov games, Doepke and Townsend [12] developed recursive algorithms to solve optimal contracts in dynamic principal-player scenarios with concealed states and hidden actions, demonstrating the feasibility of employing recursive techniques to derive optimal contracts. Their work sheds light on effective strategies for managing information asymmetry and strategic behavior and contributes to the broader research on optimal contracting in the presence of moral hazard concerns.

Chen et al. [13] devised a framework addressing moral hazard concerns encompassing both observable and unobservable actions. Their paradigm offers insights into designing incentive mechanisms that account for the complexities arising from hidden actions, therefore enhancing our understanding of how to mitigate moral hazard in various economic contexts. These research contributions underscore the importance of dynamic modeling and algorithmic approaches in analyzing principal-player interactions with hidden actions. They provide valuable guidance for designing effective incentive mechanisms and managing moral hazard concerns in dynamic and uncertain environments.

The paper addresses real-world situations where players are unable to observe each other's actions due to cost or impracticality. It offers insights into dynamic modeling, optimal contract design, and algorithmic solutions for managing strategic interactions in such environments. These findings hold practical implications for designing effective incentive mechanisms and decision-making strategies in settings where player actions are concealed.

This work introduces a dynamic mechanism design tailored for uncertain environments where parties encounter constraints due to the inability to observe players' actions. It proposes incentive schemes to tackle these constraints within a multiplayer Markov game framework, where outcomes are influenced by both the payoff state and players' actions. The involved parties are bound by either moral hazard or adverse selection, complicating decision-making. To maximize expected returns, the mechanism incentivizes truthful state revelation, aligning players' incentives with desired outcomes. The paper establishes essential relationships to determine optimal variables, including the mechanism, methods, and distribution vector. By addressing the challenges posed by hidden actions and information asymmetry, the proposed framework offers valuable insights for designing effective incentive mechanisms in complex, uncertain environments, ultimately enhancing decision-making processes and strategic interactions.

A numerical example concerning oligopolies emphasizes the effectiveness of our proposed method. By applying our approach to a specific scenario, we illustrate its practical utility and efficacy in achieving desirable outcomes. Through numerical analysis and simulation, we offer concrete evidence of the method's effectiveness in addressing real-world challenges. This example validates our approach, demonstrating its ability to deliver optimal solutions and inform its potential applications in various contexts.

The paper is structured into distinct sections to systematically address dynamic mechanism design in uncertain environments. In Section 2, the formulation of the moral hazard approach is explored, providing insights into theoretical foundations and presenting a model tailored to address this challenge. Section 3 shifts focus to mechanism design and the recovery of variables of interest, elaborating on strategies employed to mitigate the

effects of moral hazard through effective incentive mechanisms. Convergence analysis is conducted in Section 4, rigorously examining the stability and convergence properties of the proposed approach. Section 5 utilizes a numerical example to illustrate the practical utility and efficacy of the proposed method, offering concrete evidence of its effectiveness in real-world scenarios. Finally, Section 6 concludes with remarks and outlines for future research, encapsulating key findings and suggesting potential areas for further exploration and refinement of the dynamic mechanism design framework.

2. Repeated Game and Moral Hazard

The main properties of the game, considered below, are as follows:

1. The player l is informed privately of his type θ for each time $t \geq 1$ from the finite set Θ^l , and takes an action (make decisions) a_t^l from the finite set A^l [14,15];
2. all players, being in states $\Theta = \times_{l \in \mathcal{N}} \Theta^l$, $\mathcal{N} := \{1, \dots, n\}$, take actions simultaneously from the set $A = \times_{l \in \mathcal{N}} A^l$;
3. each player has a valuation function $v^l(a_t^l, \theta_t^l) \geq 0$ (utility function) that establishes current utility value for the player $l \in \mathcal{N}$.
4. the state dynamics of player l is defined by the transition distribution

$$p^l(\theta_{t+1}^l | \theta_t^l, a_t^l, \theta_{t-1}^l, a_{t-1}^l, \dots, \theta_0^l, a_0^l) = p^l(\theta_{t+1}^l | \theta_t^l, a_t^l), \tag{1}$$

that is the probability to move from type θ_t^l to type θ_{t+1}^l applying action a_t^l .

5. denoting by $\Delta(\Theta^l)$ the set of state distribution $\mu_t^l(\theta_t^l)$ over Θ^l , namely $\mu_t^l(\theta_t^l) \in \Delta(\Theta^l)$, let us suppose that the state dynamics of the players are mutually independent, each chain is ergodic, and μ^l is its unique invariant distribution, i.e.,

$$\mu_t^l(\theta_t^l) \xrightarrow{t \rightarrow \infty} \mu^l(\theta^l).$$

We associate each player's message m_t^l with its type, where $m_t^l \in M^l \subseteq \Theta^l$, and denote the overall message as $m_t = (m_t^1, \dots, m_t^n)$. Here, $M^l \subseteq \Theta^l$ represents the set of messages such that at each time t , player l sends a message m_t^l to all other players, who respond with $a_t = (a_t^1, a_t^2, \dots, a_t^n)$.

Remark 1. In partially observable Markov processes, the set of observable states is external, distinct from Θ^l . Conversely, in Bayes–Markov processes, the set of observable states is internal, corresponding to the subset M^l . This distinction highlights the difference in how information is perceived and processed by players in different types of Markov processes, impacting their decision-making and strategies within the game.

Reward Function

The utility function $v^l(a_t^l, \theta_t^l, m_t^l)$ captures the losses incurred by player l when applying action $a_t^l \in A^l$ based on the message $m_t^l \in M^l$, generated in the type (state) $\theta_t^l \in \Theta^l$. This function quantifies the impact of player l 's actions on their overall performance, considering both the current state of the system and the message conveyed. By incorporating these factors, the utility function provides valuable insights into player l 's decision-making process and aids in optimizing their strategies within the game.

A mechanism φ in the above system is the conditional distribution $\varphi(a|m)$ of possible joint actions $a \in A = \otimes_{l \in \mathcal{N}} A^l$ that respond to the joint message $m \in M = \otimes_{l \in \mathcal{N}} M^l$.

Denote by Φ_{adm} the set of possible mechanisms:

$$\Phi_{adm} = \left\{ \varphi(\alpha_t | m_t) \geq 0 \mid \sum_{\alpha_t \in A} \varphi(\alpha_t | m_t) = 1, m_t \in M \right\}. \tag{2}$$

A (behavioral) strategy $\sigma^l(m_t^l|\theta_t^l)$ for player l is given by $\sigma^l : \Theta^l \rightarrow M^l$, which associates the given type θ_t^l with the message m_t^l .

The set of feasible policies is given by

$$\mathcal{S}_{adm}^l = \left\{ \sigma^l(m_t^l|\theta_t^l) \geq 0 \mid \sum_{m_t^l \in \Theta^l} \sigma^l(m_t^l|\theta_t^l) = 1, \theta_t^l \in \Theta^l \right\}. \tag{3}$$

and

$$\mathcal{S}_{adm} = \mathcal{S}_{adm}^1 \times \dots \times \mathcal{S}_{adm}^n.$$

The relationship $\zeta^l(a_t^l|\alpha_t^l)$ for player l is given by $\zeta^l : \mathcal{A}^l \rightarrow \Delta(\mathcal{A}^l)$. The set of actions that are feasible is defined by

$$\mathcal{E}_{adm}^l = \left\{ \zeta^l(a_t^l|\alpha_t^l) \geq 0 \mid \sum_{a_t^l \in \mathcal{A}_t^l} \zeta^l(a_t^l|\alpha_t^l) = 1, \alpha_t^l \in \mathcal{A}_t^l \right\}. \tag{4}$$

Remark 2. $\zeta^l(a_t^l|\alpha_t^l)$ represents the likelihood with which a player l believes that α_t^l is an action a_t^l .

The average reward $U^l(\varphi, \sigma)$ (in the stationary ergodic regime) of the player l is the summed utility $W^l(a_t^l, \theta_t^l, m_t^l)$ given by employing the mechanism φ , the strategy σ and $\zeta^l(a_t^l|\alpha_t^l)$:

$$U^l(\varphi, \sigma, \zeta) = \sum_{\theta_t^l \in \Theta^l} \sum_{m_t^l \in M} \sum_{a_t^l \in \mathcal{A}} W^l(a_t^l, \theta_t^l, m_t^l) \prod_{i \in \mathcal{N}} \varphi(\alpha_i|m_t) \sigma^i(m_t^i|\theta_t^i) \zeta^i(a_t^i|\alpha_t^i) \mu^i(\theta_t^i), \tag{5}$$

where

$$W^l(a_t^l, \theta_t^l, m_t^l) = \sum_{\theta_{t+1}^l \in \Theta^l} v^l(a_t^l, \theta_t^l, m_t^l) p^l(\theta_{t+1}^l|\theta_t^l, a_t^l),$$

The following series of events occurs in time t and player $l \in \mathcal{N}$ given a mechanism φ :

1. Every player privately observes his current type θ_t^l obtained from $\mu^l(\theta_t^l)$.
2. Every player sends a message $m_t^l \in M_t^l$ according to $\sigma^l(m_t^l|\theta_t^l)$.
3. An alternative $\alpha_t \in A_t$ is selected considering $\varphi(\alpha_t|m_t) \in \Phi_{adm}$.
4. Every player selects an $a_t \in A_t$ given $\alpha_t^l \in \mathcal{A}_t^l$ and $\zeta^l(a_t^l|\alpha_t^l)$.
5. The allocation is realized, and players obtain a reward.
6. At last, θ_{t+1}^l is obtained from $p^l(\theta_{t+1}^l|\theta_t^l, a_t^l)$ given θ_t^l and a_t^l .

3. Mechanism and Equilibrium

Following [14,15] define the extended variable

$$c^l(m_t^l \theta_t^l a_t^l \alpha_t^l) = \varphi(\alpha_t|m_t) \sigma^l(m_t^l|\theta_t^l) \zeta^l(a_t^l|\alpha_t^l) \mu^l(\theta_t^l) \in C_{adm}^l \tag{6}$$

with

$$C_{adm}^l := \left\{ c^l(m_t^l \theta_t^l a_t^l \alpha_t^l) \geq 0 \mid \sum_{\theta_t^l \in \Theta^l} \sum_{m_t^l \in \Theta^l} \sum_{a_t^l \in \mathcal{A}_t^l} \sum_{\alpha_t^l \in \mathcal{A}_t^l} c^l(m_t^l \theta_t^l a_t^l \alpha_t^l) = 1, \right. \\ \left. \sum_{m_t^l \in \Theta^l} \sum_{a_t^l \in \mathcal{A}_t^l} \sum_{\alpha_t^l \in \mathcal{A}_t^l} c^l(m_t^l \theta_t^l a_t^l \alpha_t^l) = \mu^l(\theta_t^l) > 0, \right. \\ \left. \sum_{m_t^l \in \Theta^l} \sum_{a_t^l \in \mathcal{A}_t^l} \sum_{\theta_t^l \in \Theta^l} \sum_{\theta_{t+1}^l \in \Theta_{t+1}^l} \sum_{\alpha_t^l \in \mathcal{A}_t^l} [\delta_{\theta_t^l \theta_{t+1}^l} - p^l(\theta_{t+1}^l|\theta_t^l, a_t^l)] c^l(m_t^l \theta_t^l a_t^l \alpha_t^l) = 0 \right\}, \tag{7}$$

where $\delta_{\tilde{\theta}^l, \theta^l}$ is Kroneker's symbol, $\theta_{t+1}^l \in \Theta_t^l$ and $c(\theta_t^l m_t^l a_t^l \alpha_t^l) = (c^1(\theta_t^1 m_t^1 a_t^1 \alpha_t^1), \dots, c^n(\theta_t^n m_t^n a_t^n \alpha_t^n))$. Notice that

$$\sum_{\alpha_t^l \in A_t} \varphi(\alpha_t | m_t) = 1, \quad \sum_{m_t^l \in \Theta_t} \sigma^l(m_t^l | \theta_t^l) = 1, \quad \sum_{\theta_t^l \in \Theta_t^l} \mu^l(\theta_t^l) = 1, \quad \sum_{a_t^l \in A_t^l} \xi^l(a_t^l | \alpha_t^l) = 1$$

such that $c^l \in S^l$ satisfies

$$S^l := \left\{ \begin{aligned} &c^l(m_t^l \theta_t^l a_t^l \alpha_t^l) \geq 0 \left| \begin{aligned} &\sum_{\theta_t^l \in \Theta_t^l} \sum_{m_t^l \in \Theta_t^l} \sum_{a_t^l \in A_t^l} \sum_{\alpha_t^l \in A_t^l} c^l(m_t^l \theta_t^l a_t^l \alpha_t^l) = 1, \\ &\sum_{m_t^l \in \Theta_t^l} \sum_{a_t^l \in A_t^l} \sum_{\alpha_t^l \in A_t^l} c^l(m_t^l \theta_t^l a_t^l \alpha_t^l) = \mu^l(\theta_t^l) > 0 \end{aligned} \right. \end{aligned} \right. \quad (8)$$

The individual aim (5) in c -variables looks as

$$U^l(\varphi, \sigma, \xi) = \sum_{l \in \mathcal{N}} \bar{U}^l(c) \rightarrow \max_{c \in \mathcal{C}_{adm}} \quad (9)$$

$$\bar{U}^l(c) = \sum_{\theta_t^l \in \Theta_t^l} \sum_{m_t^l \in \Theta_t^l} \sum_{a_t^l \in A_t^l} \sum_{\alpha_t^l \in A_t^l} W^l(a_t^l, \theta_t^l, m_t^l) \prod_{i \in \mathcal{N}} c^i(m_t^i \theta_t^i a_t^i \alpha_t^i) \quad (10)$$

where $c^l(m_t^l \theta_t^l a_t^l \alpha_t^l)$ is given by (6) and $\mathcal{C}_{adm} = \times_l \mathcal{C}_{adm}^l$. The variable c^{l*} denotes the solution of the problem given in Equation (9).

Find a mechanism $\varphi(\alpha_t | m_t)$, the Bayesian strategies σ^* and the action kernel ξ^* , that realizes the rule given by and solves the following nonlinear programming problem

$$(\varphi^*, \sigma^*, \xi^*) = \arg \max_{\varphi \in \Phi_{adm}, l \in \mathcal{N}} \sum_{l \in \mathcal{N}} U^l(\varphi, \sigma^*, \xi^*). \quad (11)$$

where for a given mechanism $\varphi(a_t | m_t)$ the strategies σ_φ^* and ξ_φ^* fulfills the *Bayesian–Nash equilibrium* satisfying

$$U^l(\varphi, \sigma_\varphi^*, \xi_\varphi^*) \geq U^l(\varphi, \sigma^l, \sigma^{-l*}, \xi^l, \xi^{-l*}). \quad (12)$$

such that φ is unique and $\sigma^* = (\sigma^{1*}, \dots, \sigma^{n*})$ is known as the *Bayesian–Nash equilibrium* where $\sigma^{-l*} = (\sigma^{1*}, \dots, \sigma^{l-1*}, \sigma^{l+1*}, \dots, \sigma^{n*})$ and $\xi^{-l*} = (\xi^{1*}, \dots, \xi^{l-1*}, \xi^{l+1*}, \dots, \xi^{n*})$.

The following Lemma explains how we can compute the mechanism $\varphi^*(\alpha_t | m_t)$ given $c^{l*}(\theta_t^l m_t^l a_t^l \alpha_t^l)$,

Let us assume that (9) is solved, then the variable $\varphi^*(\alpha_t | m_t)$ can be recovered from $c^{l*}(\theta_t^l m_t^l a_t^l \alpha_t^l)$ as follows:

$$\varphi^*(\alpha_t | m_t) = \frac{\sum_{l \in \mathcal{N}} \sum_{\theta_t^l \in \Theta_t^l} \sum_{a_t^l \in A_t^l} c^{l*}(\theta_t^l m_t^l a_t^l \alpha_t^l)}{\sum_{l \in \mathcal{N}} \sum_{\theta_t^l \in \Theta_t^l} \sum_{a_t^l \in A_t^l} \sum_{\gamma_t^l \in A_t^l} c^{l*}(\theta_t^l m_t^l a_t^l \gamma_t^l)} \quad (13)$$

The variables can be retrieved using the formulas shown below.

To recover $\sigma^{l*}(m_t^l | \theta_t^l)$ note that

$$\sigma^{l*}(m_t^l | \theta_t^l) = \frac{\sum_{a_t^l \in A_t^l} \sum_{\alpha_t^l \in A_t^l} c^{l*}(\theta_t^l m_t^l a_t^l \alpha_t^l)}{\sum_{q_t^l \in \Theta_t^l} \sum_{a_t^l \in A_t^l} \sum_{\alpha_t^l \in A_t^l} c^{l*}(\theta_t^l q_t^l a_t^l \alpha_t^l)}. \quad (14)$$

For the distribution $\bar{\mu}^{l*}(m_t^l)$ the resulting relation is given by

$$\bar{\mu}^{l*}(m_t^l) = \sum_{\theta_t^l \in \Theta_t^l} \sum_{a_t^l \in A_t^l} \sum_{\alpha_t^l \in \mathcal{A}_t^l} c^{l*}(\theta_t^l m_t^l a_t^l \alpha_t^l), \quad l \in \mathcal{N} \tag{15}$$

The actions observer $\zeta_{\alpha|a}^*$ can be recovered employing

$$\zeta^{l*}(a_t^l | \alpha_t^l) = \frac{\sum_{\theta_t^l \in \Theta_t^l} \sum_{m_t^l \in \Theta_t^l} c^{l*}(\theta_t^l m_t^l a_t^l \alpha_t^l)}{\sum_{\theta_t^l \in \Theta_t^l} \sum_{m_t^l \in \Theta_t^l} \sum_{\gamma_t^l \in A_t^l} c^{l*}(\theta_t^l m_t^l a_t^l \gamma_t^l)}. \tag{16}$$

If the solutions of the problem (9) given by $\sigma^{l*}(m_t^l | \theta_t^l)$, $\varphi^*(\alpha_t | m_t)$ and $\zeta^{l*}(a_t^l | \alpha_t^l)$ comply with the Bayesian–Nash equilibrium presented in Equation (12) for all $l \in \mathcal{N}$, then variables $c^{l*}(\theta_t^l m_t^l a_t^l \alpha_t^l)$ fulfill the ergodicity constraints given by:

$$\sum_{\theta_t^l \in \Theta_t^l} \sum_{m_t^l \in \Theta_t^l} \sum_{a_t^l \in A_t^l} \sum_{\alpha_t^l \in \mathcal{A}_t^l} [\delta_{\theta_t^l \theta_{t+1}^l} - p^l(\theta_{t+1}^l | \theta_t^l a_t^l)] c^{l*}(\theta_t^l m_t^l a_t^l \alpha_t^l), \quad \theta_{t+1}^l \in \Theta_t^l, \tag{17}$$

Lemma 1. *The mechanism $\varphi^*(\alpha_t | m_t)$, the actions kernel $\zeta^{l*}(a_t^l | \alpha_t^l)$ and the strategies $\sigma^{l*}(m_t^l | \theta_t^l)$ coincide with the Bayesian–Nash equilibrium presented in Equation (12).*

Proof. It results straightforward from:

$$\begin{aligned} & \max_{c \in C_{adm}} \bar{U}(c) = \bar{U}(c^*) = \sum_{l \in \mathcal{N}} \bar{U}^l(c^*) = \sum_{l \in \mathcal{N}} U^l(\varphi^*, \sigma_{\varphi^*}^*, \zeta_{\varphi^*}^*) = \\ & \sum_{l \in \mathcal{N}} \left(\sum_{\theta_t^l \in \Theta_t^l} \sum_{m_t^l \in \Theta_t^l} \sum_{a_t^l \in A_t^l} \sum_{\alpha_t^l \in \mathcal{A}_t^l} W^l(a_t^l, \theta_t^l, m_t^l) (\varphi^*(\alpha_t | m_t))^n \sigma^{l*}(m_t^l | \theta_t^{l*}) \right. \\ & \quad \left. \zeta^{l*}(a_t^l | \alpha_t^{l*}) \mu^{l*}(\theta_t^{l*}) \prod_{i \neq l \in \mathcal{N}} \sigma^{i*}(m_t^{i*} | \theta_t^{i*}) \zeta^{i*}(\alpha_t^{i*} | a_t^{i*}) \mu^{i*}(\theta_t^{i*}) \right) = \\ & \sum_{l \in \mathcal{N}} \left(\max_{\sigma^l \in \mathcal{S}_{adm}^l, \zeta^l \in \Xi_{adm}^l} \sum_{m_t^l \in \Theta_t^l} \sum_{\alpha_t^l \in \mathcal{A}_t^l} (\varphi^*(\alpha_t | m_t))^n \sum_{\theta_t^l \in \Theta_t^l} \sum_{a_t^l \in A_t^l} W^l(a_t^l, \theta_t^l, m_t^l) \sigma^l(m_t^l | \theta_t^l) \right. \\ & \quad \left. \zeta^l(\alpha_t^l | a_t^l) \mu^l(\theta_t^l) \prod_{i \neq l \in \mathcal{N}} \sigma^{i*}(m_t^{i*} | \theta_t^{i*}) \zeta^{i*}(\alpha_t^{i*} | a_t^{i*}) \mu^{i*}(\theta_t^{i*}) \right) \\ & \geq \sum_{l \in \mathcal{N}} \left(\sum_{m_t^l \in \Theta_t^l} \sum_{\alpha_t^l \in \mathcal{A}_t^l} (\varphi^*(\alpha_t | m_t))^n \sum_{\theta_t^l \in \Theta_t^l} \sum_{a_t^l \in A_t^l} W^l(a_t^l, \theta_t^l, m_t^l) \sigma^l(m_t^l | \theta_t^l) \right. \\ & \quad \left. \zeta^l(\alpha_t^l | a_t^l) \mu^l(\theta_t^l) \prod_{i \neq l \in \mathcal{N}} \sigma^{i*}(m_t^{i*} | \theta_t^{i*}) \zeta^{i*}(\alpha_t^{i*} | a_t^{i*}) \mu^{i*}(\theta_t^{i*}) \right) = \\ & \quad \sum_{l \in \mathcal{N}} U^l(\varphi, \sigma^l, \sigma^{-l*}, \zeta^l, \zeta^{-l*}) \end{aligned}$$

As a result,

$$\sum_{l \in \mathcal{N}} \left(U^l(\varphi^*, \sigma_{\varphi^*}^*, \zeta_{\varphi^*}^*) - U^l(\varphi^*, \sigma^l, \sigma^{-l*}, \zeta^l, \zeta^{-l*}) \right) \geq 0. \tag{18}$$

Given that Equation (18) is valid for all the feasible strategies σ and ζ , it is satisfied for $\sigma^j = \sigma^{j*}$ and $\zeta^j = \zeta^{j*}$ for $j \neq l$, with

$$U^l(\varphi^*, \sigma_{\varphi^*}^*, \zeta_{\varphi^*}^*) - U^l(\varphi^*, \sigma^l, \sigma^{-l*}, \zeta^l, \zeta^{-l*}) \geq 0, \tag{19}$$

that satisfies Equation (12) for $\varphi = \varphi^*$. \square

4. Convergence Analysis

In our study, we illustrate the convergence of our proposed strategy to a Bayesian–Nash equilibrium, as documented in previous works [16–18]. To ensure this convergence, we incorporate a regularization parameter into our strategy. This parameter acts as a stabilizing factor, guiding the system towards a single equilibrium point. Through empirical analysis and simulations, we validate the effectiveness of our approach in achieving convergence to the desired equilibrium, highlighting the importance of the regularization parameter in enhancing stability and optimality in strategic interactions.

The game comprises a set $\mathcal{N} = 1, \dots, n$ of players indexed by $l \in \mathcal{N}$. Each player l 's joint strategy variable is denoted as

$$c^l(m_t^l, \theta_t^l, a_t^l, \alpha_t^l) = \varphi(\alpha_t | m_t) \sigma^l(m_t^l | \theta_t^l) \xi^l(a_t^l | \alpha_t^l) \mu^l(\theta_t^l).$$

This variable encapsulates the interplay of various components, including message, type, action, and belief functions, shaping player l 's strategic decisions within the game.

In the game, variables $x^l \in \mathcal{X}^l$ are considered, where \mathcal{X}^l represents a convex and compact set. Here, $x^l := \text{col}(c^l(m_t^l, \theta_t^l, a_t^l, \alpha_t^l))$, where col denotes the column operator. This formulation aggregates player l 's joint strategy variable into a column vector, facilitating mathematical representation and analysis of the game dynamics. Let $x = (x^1, \dots, x^n)^\top \in \mathcal{X}$ be the joint strategy of the players and let $x^{-l} := (x^1, \dots, x^{l-1}, x^{l+1}, \dots, x^n)^\top \in \mathcal{X}^{-l}$ be the strategy of the complement of the players. Let us consider a Nash equilibrium problem with n players and denote by $x = (x^l, x^{-l}) \in \mathbb{R}^n$ the vector representing the x -th player's strategy where $x \in \mathcal{X}_{adm}$ and $\mathcal{X}_{adm} = \mathcal{X}_{adm}^l \times \mathcal{X}_{adm}^{-l}$ is a bounded set given by

$$\mathcal{X}_{adm} := \{x \in \mathbb{R}^n : x \geq 0, A_{eq}x = b_{eq} \in \mathbb{R}^{m_0}, A_{ineq}x \leq b_{ineq} \in \mathbb{R}^{m_1}\}.$$

Let $g^l : \mathbb{R}^n \rightarrow \mathbb{R}$ represent the l -th player's reward function (or cost function). Players aim to reach a non-cooperative equilibrium by finding a strategy $x^* = (x^{1*}, \dots, x^{n*})$ where for any x^l and any $l \in \mathcal{N}$:

$$\tilde{U}(x) := \sum_{l \in \mathcal{N}} \left[\left(\max_{x^l \in \mathcal{X}^l} g^l(x^l, x^{-l}) \right) - g^l(x^l, x^{-l}) \right]. \tag{20}$$

Let

$$\tilde{U}(x) := \sum_{l \in \mathcal{N}} \left[g^l(\bar{x}^l, x^{-l}) - g^l(x^l, x^{-l}) \right], \tag{21}$$

and

$$\bar{x}^l \in \text{Arg max}_{x^l \in \mathcal{X}^l} g^l(x^l, x^{-l}) \tag{22}$$

where

$$g^l(\bar{x}^l, x^{-l}) - g^l(x^l, x^{-l}) \geq 0 \tag{23}$$

for any $x^l \in \mathcal{X}^l$ and $l \in \mathcal{N}$. The vector $x^* \in \mathcal{X}$ is a Nash equilibrium if

$$x^* \in \text{Arg max}_{x \in \mathcal{X}} \{ \tilde{U}(x) \}. \tag{24}$$

Lagrange Regularization Approach

In our analysis, we denote by $\mathcal{X}^* \subseteq \mathcal{X}_{adm}$ the set encompassing all potential solutions of the game [19,20]. To ensure the attainment of a unique solution within this set, we employ the Lagrange function. This mathematical tool enables the incorporation of constraints into the optimization framework by integrating the objective function with penalty terms for constraint violations. By doing so, the Lagrange function facilitates the identification of an optimal solution that simultaneously satisfies both the objective and constraints. Through

the careful utilization of the Lagrange function, we navigate the complexities inherent in the game and ensure the achievement of a unique solution within the designated solution space. This approach allows us to address the intricacies of the problem and obtain a solution that optimally balances the competing objectives and constraints involved. To reach a unique solution, we consider the Lagrange function given by

$$\mathbb{L}(x, \lambda_{eq}, \lambda_{ineq}) := -\omega \tilde{U}(x) - \lambda_{eq}^T (A_{eq}x - b_{eq}) - \lambda_{ineq}^T (A_{ineq}x - b_{ineq}) - \frac{\delta}{2} \left(\|x\|^2 - \|\lambda_{eq}^T\|^2 - \|\lambda_{ineq}^T\|^2 \right) \tag{25}$$

where the parameters ω and δ are positive, the Lagrange vector-multipliers $\lambda_{ineq} \in \mathbb{R}^{m_1}$ are non-negative and the components of $\lambda_{eq} \in \mathbb{R}^{m_0}$ may have any sign. As a result, the nonlinear problem

$$\mathbb{L}_{\omega, \delta}(x, \lambda_{eq}, \lambda_{ineq}) \rightarrow \min_{x \in \mathcal{X}_{adm}} \max_{\lambda_{eq}, \lambda_{ineq} \geq 0} \tag{26}$$

has a unique saddle point on z since the optimized Equation (25) is *strongly convex* if the parameters ω and $\delta > 0$ provide the condition

$$\frac{\partial^2}{\partial x \partial x^T} \mathbb{L}_{\omega, \delta}(x, \lambda_{eq}, \lambda_{ineq}) > 0 \forall x \in \mathcal{X}_{adm} \subset \mathbb{R}^n \tag{27}$$

and it is strongly concave on the Lagrange multipliers $\lambda_{eq}, \lambda_{ineq}$ for any $\delta > 0$. Given the properties of Equation (25) it has a unique saddle point

$$\left(x^*(\delta), \lambda_{eq}^*(\omega, \delta), \lambda_{ineq}^*(\omega, \delta) \right) \tag{28}$$

for which the following inequalities hold:

$$\mathbb{L}_{\omega, \delta}(x(\delta), \lambda_{eq}^*(\omega, \delta), \lambda_{ineq}^*(\omega, \delta)) \geq \mathbb{L}_{\omega, \delta}(x^*(\delta), \lambda_{eq}^*(\omega, \delta), \lambda_{ineq}^*(\omega, \delta)) \geq \mathbb{L}_{\omega, \delta}(x^*(\delta), \lambda_{eq}(\omega, \delta), \lambda_{ineq}(\omega, \delta)) \tag{29}$$

for any $\lambda_{eq}, \lambda_{ineq}$ with non-negative components and any $x \in \mathbb{R}^n$.

Theorem 1. *The nonlinear programming problem*

$$\mathbb{L}_{\omega, \delta}(x, \lambda_{eq}, \lambda_{ineq}) \rightarrow \min_{x \in \mathcal{X}_{adm}} \max_{\lambda_{eq}, \lambda_{ineq} \geq 0}$$

has a unique equilibrium point at x .

Remark 3. *This theorem asserts that the given nonlinear programming problem possesses a single equilibrium point located at x , indicating a stable solution within the optimization framework.*

Proof. Considering that the functional is minimizing, we prove that the Hessian matrix $\mathcal{H} := \frac{\partial^2}{\partial x \partial x^T} \mathbb{L}_{\omega, \delta}(x, \lambda_{eq}, \lambda_{ineq})$ is strictly positive definite, $\mathcal{H} > 0$, for all $x \in \mathbb{R}^n$ and for some $\omega, \delta > 0$. Developing \mathcal{H} one has that

$$\frac{\partial^2}{\partial x^2} \mathbb{L}_{\omega, \delta}(x, \lambda_{eq}, \lambda_{ineq}) = \omega \frac{\partial^2}{\partial x^2} U(x) + \delta I_{n \times n} \geq \delta \left(1 + \frac{\omega}{\delta} \zeta \right) I_{n \times n} > 0 \forall \delta > \omega |\zeta|$$

$$\zeta := \min_{x \in \mathcal{X}_{adm}} \zeta_{\min} \left(\frac{\partial^2}{\partial x^2} U(x) \right)$$

where ζ is an eigenvalue that fulfills the condition $\mathcal{H} > 0$ if $\delta > \omega |\zeta|$. Then, the Lagrange in Equation (25) has a unique minimal point x^* . \square

Theorem 2. Under the assumption that the bounded set \mathcal{X}^* of all solutions of the game is not empty and there exists an equilibrium point $\tilde{x} \in \mathcal{X}_{adm}$ such that

$$A_{ineq}\tilde{x} < b_{ineq} \tag{30}$$

that is, Slater’s condition holds; furthermore, that the parameters ω and δ are time-varying, i.e.,

$$\omega = \omega_i, \delta = \delta_i \ (i = 0, 1, 2, \dots)$$

such that

$$0 < \omega_i \downarrow 0, \frac{\omega_i}{\delta_i} \downarrow 0 \text{ when } i \rightarrow \infty, \tag{31}$$

one obtains

$$\left. \begin{aligned} x_i^* &:= x^*(\omega_i, \delta_i) \xrightarrow{i \rightarrow \infty} x^{**} \\ \lambda_{eq}^*(\omega_i, \delta_i) &\xrightarrow{i \rightarrow \infty} \lambda_{eq}^{**} \\ \lambda_{ineq}^*(\omega_i, \delta_i) &\xrightarrow{i \rightarrow \infty} \lambda_{ineq}^{**} \end{aligned} \right\} \tag{32}$$

where $x^{**} \in \mathcal{X}^*$, $(\lambda_{eq}^{**}, \lambda_{ineq}^{**}) \in \Xi^*$ define the solution of the game with the minimal norm given by

$$\left. \begin{aligned} \|x^{**}\|^2 + \|\lambda_{eq}^{**}\|^2 + \|\lambda_{ineq}^{**}\|^2 &\leq \|x^*\|^2 + \|\lambda_{eq}^*\|^2 + \|\lambda_{ineq}^*\|^2 \\ \text{for all } x^* \in \mathcal{X}^*, (\lambda_{eq}^*, \lambda_{ineq}^*) &\in \Xi^* \end{aligned} \right\} \tag{33}$$

Remark 4. Theorem 2 presents conditions under which a unique solution of the game can be attained. First, we assume that the bounded set \mathcal{X}^* of all solutions of the game is not empty and there exists an equilibrium point $\tilde{x} \in \mathcal{X}_{adm}$ such that Slater’s condition (30) holds. This condition ensures the existence of a feasible point in the interior of the inequality constraints. Additionally, the parameters ω and δ are considered to be time-varying, satisfying (31) as i tends to infinity.

Under these assumptions, the theorem asserts the convergence of the solution sequences x_i^* , λ_{eq}^* , and λ_{ineq}^* to x^{**} , λ_{eq}^{**} , and λ_{ineq}^{**} , respectively, as i tends to infinity (32). Here, $x^{**} \in \mathcal{X}^*$ and $(\lambda_{eq}^{**}, \lambda_{ineq}^{**}) \in \Xi^*$ define the solution of the game with the minimal norm (33).

In essence, Theorem 2 establishes conditions under which a unique equilibrium solution of the game can be achieved, ensuring convergence of the solution sequences to a solution with minimal norm. These conditions are crucial for guaranteeing the stability and optimality of the solution in the dynamic environment described by the nonlinear programming problem.

Proof. See Appendix A. \square

5. Numerical Example

5.1. Description of the Oligopoly Problem

In the typical structure of an oligopolistic market, only a handful of producers offer similar or identical products, resulting in homogeneous goods [21]. This market arrangement is characterized by a known demand function, with consumers playing a passive role. In this scenario, the producers themselves act as the players in the market dynamics. To address the challenges posed by oligopoly, producers often agree to split the market among themselves, recognizing that fragmenting the market further into smaller companies would be counterproductive. The price each producer receives for their product is influenced by the quantity of goods produced, and producers react to each other’s production adjustments until an equilibrium is reached. This equilibrium, known as a Nash equilibrium, occurs when no producer can improve their payoff by unilaterally changing their strategy, given the strategies of the other producers. In essence, producers in an oligopolistic market strive to balance their production decisions and pricing strategies to achieve a stable

market outcome. This strategic interaction among producers is a defining characteristic of oligopoly, shaping market dynamics and outcomes.

5.2. Resulting Values

Applying our method yields a resulting mechanism that effectively optimizes strategic interactions. This mechanism orchestrates player behaviors to converge towards desired outcomes, promoting stability and efficiency within the game environment. Through iterative refinement and adjustment, it guides players towards actions that maximize collective welfare while respecting constraints and objectives.

$$\varphi_{\alpha|m}^* = \begin{bmatrix} 0.3287 & 0.3287 & 0.3426 \\ 0.2637 & 0.2637 & 0.4727 \\ 0.2302 & 0.2302 & 0.5396 \\ 0.3111 & 0.3110 & 0.3779 \end{bmatrix}$$

The resulting behavior strategies emerge from our method, encapsulating players' optimal actions within the game. These strategies are finely tuned to navigate the complexities of strategic interactions, aiming to achieve desirable outcomes while balancing competing objectives and constraints effectively.

$$\sigma_{m|\theta}^{1*} = \begin{bmatrix} 0.1918 & 0.3711 & 0.1921 & 0.2450 \\ 0.2500 & 0.2500 & 0.2500 & 0.2500 \\ 0.2503 & 0.2499 & 0.2495 & 0.2502 \\ 0.2423 & 0.2422 & 0.2733 & 0.2422 \end{bmatrix} \quad \sigma_{m|\theta}^{2*} = \begin{bmatrix} 0.1124 & 0.3184 & 0.4568 & 0.1124 \\ 0.2650 & 0.2416 & 0.2383 & 0.2551 \\ 0.2597 & 0.2384 & 0.2389 & 0.2630 \\ 0.2413 & 0.2466 & 0.2715 & 0.2407 \end{bmatrix}$$

$$\sigma_{m|\theta}^{3*} = \begin{bmatrix} 0.1757 & 0.1760 & 0.4518 & 0.1964 \\ 0.2425 & 0.2418 & 0.2746 & 0.2411 \\ 0.2436 & 0.2436 & 0.2436 & 0.2693 \\ 0.2500 & 0.2500 & 0.2500 & 0.2500 \end{bmatrix}$$

The resulting action kernels represent the distribution of players' actions based on their observed information and strategic considerations. These kernels encapsulate the strategic decision-making process within the game, guiding players towards actions that maximize their utility while considering the actions of others and the overall game dynamics.

$$\zeta_{a|\alpha}^{1*} = \begin{bmatrix} 0.3286 & 0.3286 & 0.3427 \\ 0.3333 & 0.3333 & 0.3333 \\ 0.2324 & 0.2324 & 0.5352 \end{bmatrix} \quad \zeta_{a|\alpha}^{2*} = \begin{bmatrix} 0.3254 & 0.3254 & 0.3491 \\ 0.3172 & 0.3172 & 0.3656 \\ 0.1760 & 0.1760 & 0.6480 \end{bmatrix}$$

$$\zeta_{a|\alpha}^{3*} = \begin{bmatrix} 0.3333 & 0.3333 & 0.3333 \\ 0.3292 & 0.3292 & 0.3416 \\ 0.2126 & 0.2126 & 0.5748 \end{bmatrix}$$

The distribution vectors outline the probabilistic allocation of resources or outcomes within the game. These vectors represent the distribution of strategic variables or states among players, providing insights into the distributional dynamics and resource allocation strategies employed within the game environment.

$$\mu_{\theta}^{1*} = \begin{bmatrix} 0.4537 \\ 0.1800 \\ 0.1805 \\ 0.1858 \end{bmatrix} \quad \mu_{\theta}^{2*} = \begin{bmatrix} 0.4032 \\ 0.1974 \\ 0.2024 \\ 0.1971 \end{bmatrix} \quad \mu_{\theta}^{3*} = \begin{bmatrix} 0.4468 \\ 0.1884 \\ 0.1848 \\ 0.1800 \end{bmatrix}$$

The convergence of strategies is illustrated in Figures 1–3. These figures track the evolution of strategies over time, demonstrating their convergence towards stable equilibria. They offer valuable insights into the dynamics of strategic interactions within the game, highlighting how players' behaviors evolve and stabilize over iterations.

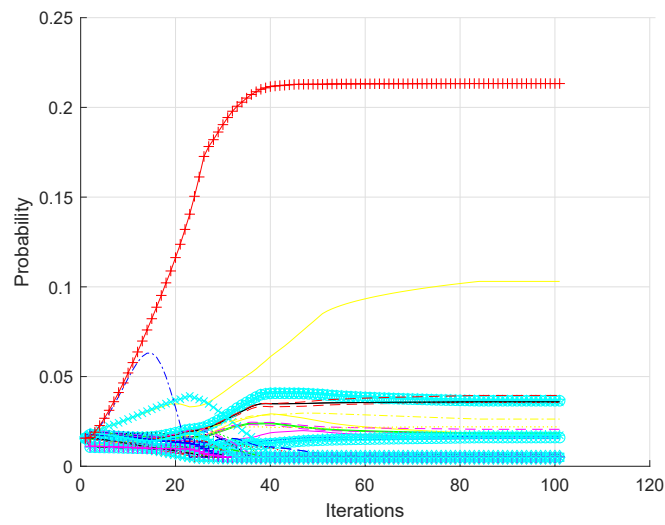


Figure 1. Convergence of the Strategies c^1 of the Player 1.

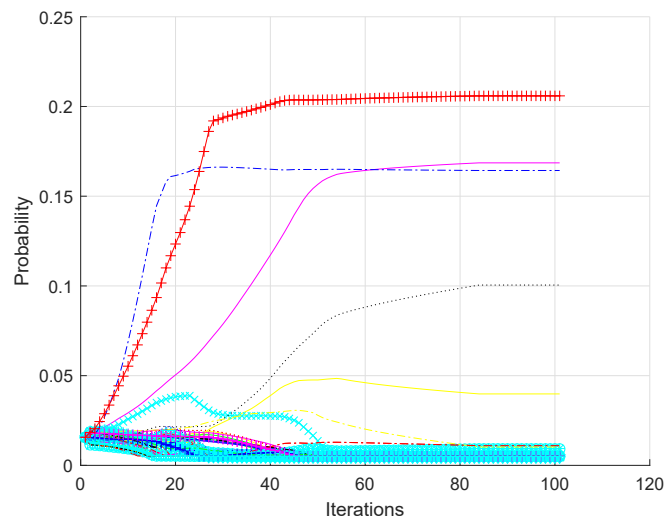


Figure 2. Convergence of the Strategies c^2 of the Player 2.

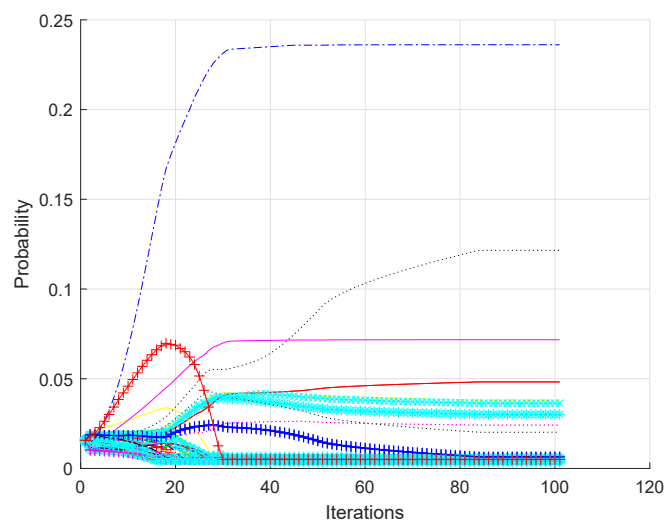


Figure 3. Convergence of the Strategies c^3 of the Player 3.

Designing a mechanism that aligns the equilibrium state with the market state, as prescribed by the welfare theorem, offers a promising approach to addressing the complexities of oligopolies with homogeneous products. The literature on this topic extensively investigates the presence of Nash equilibrium in dominant tactics, employing various mechanism designs such as strategic games. In strategic games, players may lack the motivation to provide accurate information, and they are incentivized to reveal private information to manipulate outcomes. This dynamic underscores the intricate interplay between strategic behavior and information asymmetry in oligopolistic markets. Using RL methods, this example provides a game-theoretic study of economic systems and observers. RL offers a powerful tool for analyzing complex interactions among agents in economic systems, allowing for the exploration of emergent behaviors and equilibrium states. By leveraging RL techniques, researchers can gain insights into the strategic decision-making processes of agents in oligopolistic markets, shedding light on optimal strategies and equilibrium outcomes. This approach enriches our understanding of economic dynamics and informs the design of more effective mechanisms for managing oligopolistic competition.

6. Conclusions and Future Work

The paper introduced a pioneering mechanism design tailored for uncertain environments marked by the inability to observe players' actions. Within the framework of a Markov game, players observe outcomes influenced by both the payoff state and their actions while constrained by either moral hazard or adverse selection. The proposed mechanism incentivizes players to disclose truthful information about their states, thus maximizing expected payoffs. This perspective offers valuable insights into the Bayesian literature by addressing challenges posed by hidden actions and information asymmetry in dynamic environments. Integrating concepts from game theory and Bayesian analysis, the model provides a robust framework for designing effective incentive schemes aligning players' incentives with desired outcomes. Emphasizing truthful state revelation underscores the importance of aligning incentives to mitigate adverse effects of moral hazard and adverse selection. This approach not only enhances decision-making processes but also contributes to developing more efficient and equitable systems in uncertain environments.

A numerical example centered on oligopolies underscores the effectiveness of our proposed method. Through this example, we showcase how our approach can adeptly tackle challenges within oligopolistic markets, highlighting its practical relevance and applicability in real-world scenarios. Crafting a mechanism that synchronizes the equilibrium state with the market state, as advocated by the welfare theorem, presents a promising avenue for navigating the intricacies of oligopolies with homogeneous products. Extensive literature on this subject delves into probing the existence of Nash equilibrium under dominant tactics, employing diverse mechanism designs like strategic games. In strategic games, players might lack the motivation to furnish accurate information, as they are incentivized to disclose private information to influence outcomes. This exploration underscores the challenge of aligning incentives to induce truthful information revelation, which is vital for devising effective mechanisms to address the complexities inherent in oligopolistic market dynamics.

The model's relevance extends beyond theoretical realms, offering practical implications for various real-world scenarios characterized by prevalent information asymmetry. By presenting a systematic approach to mechanism design in uncertain environments, the paper provides valuable insights for policymakers, practitioners, and researchers grappling with complex decision-making challenges in dynamic settings. The proposed mechanism design represents a significant contribution to the Bayesian literature, introducing a fresh perspective on addressing the hurdles of hidden actions and information asymmetry. Its focus on promoting truthful state revelation and maximizing expected payoffs underscores its significance and versatility across diverse contexts, rendering it a valuable tool for advancing our understanding of decision-making under uncertainty.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Proof of Theorem 2

Given that

$$(\nabla f(w), (y - w)) \leq f(y) - f(w) \text{ and } (\nabla f(w), (w - y)) \geq f(w) - f(y)$$

are true for any convex function $f(w)$ and any w and y .

For the Lagrange function given in Equation (25) at the admissible points x, λ_{eq} and λ_{ineq} and $x_i^* = x^*(\omega_i, \delta_i), \lambda_{eq,i}^* = \lambda_{eq}^*(\omega_i, \delta_i), \lambda_{ineq,i}^* = \lambda_{ineq}^*(\omega_i, \delta_i)$, then one obtains

$$\begin{aligned} & \left[x - x_i^*, \frac{\partial}{\partial x} \mathbb{L}_{\omega_i, \delta_i}(x, \lambda_{eq}, \lambda_{ineq}) \right] - \left[\lambda_{eq} - \lambda_{eq,i}^*, \frac{\partial}{\partial \lambda_{eq}} \mathbb{L}_{\omega_i, \delta_i}(x, \lambda_{eq}, \lambda_{ineq}) \right] - \\ & \left[\lambda_{ineq} - \lambda_{ineq,i}^*, \frac{\partial}{\partial \lambda_{ineq}} \mathbb{L}_{\omega_i, \delta_i}(x, \lambda_{eq}, \lambda_{ineq}) \right] \end{aligned} \tag{A1}$$

as a result one has that

$$\begin{aligned} & \left[x - x_i^*, \frac{\partial}{\partial x} \mathbb{L}_{\omega_i, \delta_i}(x, \lambda_{eq}, \lambda_{ineq}) \right] - \left[\lambda_{eq} - \lambda_{eq,i}^*, \frac{\partial}{\partial \lambda_{eq}} \mathbb{L}_{\omega_i, \delta_i}(x, \lambda_{eq}, \lambda_{ineq}) \right] - \\ & \left[\lambda_{ineq} - \lambda_{ineq,i}^*, \frac{\partial}{\partial \lambda_{ineq}} \mathbb{L}_{\omega_i, \delta_i}(x, \lambda_{eq}, \lambda_{ineq}) \right] = \mathbb{L}_{\omega_i, \delta_i}(x, \lambda_{eq,i}^*, \lambda_{ineq,i}^*) - \\ & \mathbb{L}_{\omega_i, \delta_i}(x, \lambda_{eq}, \lambda_{ineq}) + \frac{\delta_i}{2} \left(\|x - x_i^*\|^2 + \|\lambda_{eq} - \lambda_{eq,i}^*\|^2 + \|\lambda_{eq} - \lambda_{ineq,i}^*\|^2 \right) \end{aligned}$$

By applying the saddle-point condition given in Equation (29) and in view of the complementary slackness conditions

$$\left(\lambda_{ineq}^* \right)_j \left(A_{ineq}^\top x^* - b_{ineq} \right)_i = \left(\lambda_{ineq,i}^* \right)_j \left(A_{ineq} x_i^* - b_{ineq} \right)_j = 0$$

one obtains

$$\omega_i (x^* - x_i^*)^\top \frac{\partial}{\partial x} U(x^*) + \delta_i (x^* - x_i^*)^\top x^* + \delta_i \left(\lambda_{eq}^* - \lambda_{eq,i}^* \right)^\top \lambda_{eq}^* + \left(\lambda_{ineq}^* - \lambda_{ineq,i}^* \right)^\top \delta_i \lambda_{ineq}^* \geq 0$$

Dividing both sides of this inequality by δ_i and taking $\frac{\omega_i}{\delta_i} \xrightarrow{i \rightarrow \infty} 0$, the result is

$$0 \leq \limsup_{i \rightarrow \infty} \left[(x^* - x_i^*)^\top x^* + \left(\lambda_{eq}^* - \lambda_{eq,i}^* \right)^\top \lambda_{eq}^* + \left(\lambda_{ineq}^* - \lambda_{ineq,i}^* \right)^\top \lambda_{ineq}^* \right] \tag{A2}$$

This means that there necessarily exist subsequences δ_k and ω_k ($k \rightarrow \infty$) on which there exist the limits

$$\begin{aligned} x_k^* &= z^*(\omega_k, \delta_k) \rightarrow \tilde{x}^*, \lambda_{eq,k}^* = \lambda_{eq}^*(\omega_k, \delta_k) \rightarrow \tilde{\lambda}_{eq}^* \\ \lambda_{ineq,k}^* &= \lambda_{ineq}^*(\omega_k, \delta_k) \rightarrow \tilde{\lambda}_{ineq}^* \text{ as } k \rightarrow \infty \end{aligned}$$

Suppose that there exist two limit points for two different convergent subsequences, i.e., there exist the limits

$$\begin{aligned} x_{k'}^* &= x^*(\omega_{k'}, \delta_{k'}) \rightarrow \bar{x}^*, \lambda_{eq,k'}^* = \lambda_{eq}^*(\omega_{k'}, \delta_{k'}) \rightarrow \bar{\lambda}_{eq}^* \\ \lambda_{ineq,k'}^* &= \lambda_{ineq}^*(\omega_{k'}, \delta_{k'}) \rightarrow \bar{\lambda}_{ineq}^* \text{ as } k \rightarrow \infty \end{aligned}$$

Then on these subsequences one has

$$\begin{aligned} 0 &\leq (x^* - \tilde{x}^*)^\top x^* + (\lambda_{eq}^* - \tilde{\lambda}_{eq}^*)^\top \lambda_{eq}^* + (\lambda_{ineq}^* - \tilde{\lambda}_{ineq}^*)^\top \lambda_{ineq}^* \\ 0 &\leq (x^* - \tilde{x}^*)^\top x^* + (\lambda_{eq}^* - \tilde{\lambda}_{eq}^*)^\top \lambda_{eq}^* + (\lambda_{ineq}^* - \tilde{\lambda}_{ineq}^*)^\top \lambda_{ineq}^* \end{aligned}$$

From these inequalities, it follows that points $(\tilde{x}^*, \tilde{\lambda}_{eq}^*, \tilde{\lambda}_{ineq}^*)$ and $(\bar{x}^*, \bar{\lambda}_{eq}^*, \bar{\lambda}_{ineq}^*)$ correspond to the minimum point of the function

$$s(x^*, \lambda_{eq}^*, \lambda_{ineq}^*) := \frac{1}{2} \left(\|x^*\|^2 + \|\lambda_{eq}^*\|^2 + \|\lambda_{ineq}^*\|^2 \right)$$

But the function $s(x^*, \lambda_{eq}^*, \lambda_{ineq}^*)$ is strictly convex, and that gives $\tilde{x}^* = \bar{x}^*$, $\tilde{\lambda}_{eq}^* = \bar{\lambda}_{eq}^*$, $\tilde{\lambda}_{ineq}^* = \bar{\lambda}_{ineq}^*$ hence, its minimum is unique.

References

- Zhang, H.; Zenios, S. A Dynamic Principal-Agent Model with Hidden Information: Sequential Optimality through Truthful State Revelation. *Oper. Res.* **2008**, *56*, 681–696. [[CrossRef](#)]
- Bolton, P.; Dewatripont, M. *Contract Theory*; MIT Press: Cambridge, MA, USA, 2005.
- Salanie, B. *The Economics of Contracts: A Primer*; MIT Press: Cambridge, MA, USA, 2005.
- Clempner, J.B.; Poznyak, A.S. *Optimization and Games for Controllable Markov Chains: Numerical Methods with Application to Finance and Engineering*; Springer: Cham, Switzerland, 2023.
- Cvitanić, J.; Possamaï, D.; Touzi, N. Moral hazard in dynamic risk management. *Manag. Sci.* **2016**, *63*, 3147–3529. [[CrossRef](#)]
- Garrett, D.F.; Pavan, A. Managerial turnover in a changing world. *J. Political Econ.* **2012**, *120*, 879–925. [[CrossRef](#)]
- Board, S.; Meyer-ter-Vehn, M. Reputation for quality. *Econometrica* **2013**, *81*, 2381–2462.
- Halac, M.; Kartik, N.; Liu, Q. Optimal contracts for experimentation. *Rev. Econ. Stud.* **2016**, *83*, 1040–1091. [[CrossRef](#)]
- Fudenberg, D.; Holmstrom, B.; Milgrom, P. Short-term contracts and long-term agency relationships. *J. Econ. Theory* **1990**, *51*, 1–31. [[CrossRef](#)]
- Plambeck, E.; Zenios, S. Performance-Based Incentives in a Dynamic Principal-Agent Model. *Manuf. Serv. Oper. Manag.* **2000**, *2*, 240–263. [[CrossRef](#)]
- Cole, H.L.; Kocherlakota, N. Dynamic Games with Hidden Actions and Hidden States. *J. Econ. Theory* **2001**, *98*, 114–126. [[CrossRef](#)]
- Doepke, M.; Townsend, R.M. Dynamic mechanism design with hidden income and hidden actions. *J. Econ. Theory* **2006**, *126*, 235–285. [[CrossRef](#)]
- Chen, B.; Chen, Y.; Rietzke, D. Simple contracts under observable and hidden actions. *Econ. Theory* **2020**, *69*, 1023–1047. [[CrossRef](#)]
- Clempner, J.B.; Poznyak, A.S. A nucleus for Bayesian Partially Observable Markov Games: Joint observer and mechanism design. *Eng. Appl. Artif. Intell.* **2020**, *95*, 103876. [[CrossRef](#)]
- Clempner, J.B.; Poznyak, A.S. Analytical Method for Mechanism Design in Partially Observable Markov Games. *Mathematics* **2021**, *9*, 321. [[CrossRef](#)]
- Clempner, J.B.; Poznyak, A.S. Computing The Strong Nash Equilibrium for Markov Chains Games. *Appl. Math. Comput.* **2015**, *265*, 911–927. [[CrossRef](#)]
- Clempner, J.B.; Poznyak, A.S. Convergence Analysis for Pure and Stationary Strategies in Repeated Potential Games: Nash, Lyapunov and Correlated Equilibria. *Expert Syst. Appl.* **2016**, *46*, 474–484. [[CrossRef](#)]
- Clempner, J.B. On Lyapunov Game Theory Equilibrium: Static and Dynamic Approaches. *Int. Game Theory Rev.* **2018**, *20*, 1750033. [[CrossRef](#)]
- Clempner, J.B.; Poznyak, A.S. A Tikhonov regularized penalty function approach for solving polylinear programming problems. *J. Comput. Appl. Math.* **2018**, *328*, 267–286. [[CrossRef](#)]
- Clempner, J.B.; Poznyak, A.S. A Tikhonov Regularization Parameter Approach for Solving Lagrange Constrained Optimization Problems. *Eng. Optim.* **2018**, *50*, 1996–2012. [[CrossRef](#)]
- Clempner, J.B.; Poznyak, A.S. Analyzing an Optimistic Attitude for the Leader Firm in Duopoly Models: A Strong Stackelberg Equilibrium Based on a Lyapunov Game Theory Approach. *Econ. Comput. Econ. Cybern. Stud. Res.* **2016**, *4*, 41–60.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.